Forecasting of Key Performance Indicators Based on Transformer Model

Claudia Diamantini[1][a] ,Tarique Khan[1][b], Alex Mircoli[2][c] and Domenico Potena[1][d]
[1]*Department of Information Engineering, Università Politecnica delle Marche, Italy*
[2]*Department of Economic and Social Sciences, Università Politecnica delle Marche, Italy*

Keywords: Deep Learning, Neural Networks, Transformer, Organization Process Management, KPIs, Performance Indicators.

Abstract: Key performance indicators (KPIs) express the company's strategy and vision in terms of goals and enable alignment with stakeholder expectations. In business intelligence, forecasting KPIs is pivotal for strategic decision-making. For this reason, in this work we focus on forecasting KPIs. We built a transformer model architecture that outperforms conventional models like Multi-Layer Perceptrons (MLP), Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) in KPI forecasting over the Rossmann Store, supermarket 1, and 2 datasets. Our results highlight the revolutionary potential of using cutting-edge deep learning models such as the Transformer.

1 INTRODUCTION

Responding to the digital data-driven society, companies and other organizations can now obtain a vast amount of information in various formats. Employing the available data effectively can lead to modifications in an organization's processes, systems, and procedures, as continuous business enhancement is necessary. Utilizing Key Performance Indicators (KPIs) helps in maintaining high levels of performance (Tsai and Cheng, 2012). By diligently tracking and analyzing relevant metrics in real-time, it is possible to effectively pinpoint and understand limitations, evaluate the productivity of both employees and machines, establish more ambitious objectives, and successfully achieve them by progressing forward. Measuring performance allows individuals to identify any concerns regarding their performance, assess their progress toward their objectives, and provide specific instructions for resolving any issues (Horváthová et al., 2015). Forecasting KPIs which is the fundamental goal of this work, is an essential undertaking for numerous businesses and organizations.
Various types of Key Performance Indicators (KPIs), such as cost and time KPIs, can be created based on the specific situation. However, predicting KPIs can be a difficult endeavor due to the complex and ever-changing patterns that they typically reflect. Market trends, consumer behavior, seasonal variations, events, and anomalies are just a few of the internal and external factors that have an impact on these patterns. So, regular forecasting methods like statistical models, time series models, and simple machine learning models often have trouble understanding how KPIs work and how they are connected, which leads to predictions that are not accurate or reliable. A possible solution is represented by deep learning models, which provide a more sophisticated and flexible approach to KPI analysis and prediction.
Deep learning and transformer-based models have become increasingly popular because of their ability to analyze massive and diverse datasets and identify significant patterns and relationships that are important for forecasting (Emmert-Streib et al., 2020). These models are a type of artificial neural networks that comprise numerous layers of nonlinear transformations. These layers allow the models to acquire intricate and abstract representations of data.
This study presents a transformer-based model for predicting KPIs. Our methodology harnesses the benefits of different neural network layers while addressing the drawbacks of conventional forecasting techniques. We also considered other deep learning models, including MLP, LSTM, CNN, and RNNs, and compared them with the transformer model.

[a] https://orcid.org/0000-0001-8143-7615
[b] https://orcid.org/0009-0004-1756-3560
[c] https://orcid.org/0000-0001-6459-0335
[d] https://orcid.org/0000-0002-7067-5463

Diamantini, C., Khan, T., Mircoli, A. and Potena, D.
Forecasting of Key Performance Indicators Based on Transformer Model.
DOI: 10.5220/0012726500003690

In *Proceedings of the 26th International Conference on Enterprise Information Systems (ICEIS 2024) - Volume 2*, pages 280-287
ISBN: 978-989-758-692-7; ISSN: 2184-4992

We assessed their performance on diverse real-world datasets of KPIs.

The subsequent sections of this work are structured as follows: Section 2 provides a comprehensive analysis of KPI forecasting by reviewing relevant literature. Section 3 outlines the proposed methodology for predicting Key Performance Indicators (KPIs) using deep learning models and gives information about the Transformer architecture. Section 4 presents the experiments while Section 5 discusses the results. Conclusions are drawn in Section 6, where the key findings and contributions are summarized.

2 LITERATURE STUDY

Key Performance Indicators (KPIs) are crucial tools for evaluating business performance However, managing and prioritizing KPIs can be challenging, leading to research on various methods dealing with issues such as modeling, maintenance, and expressiveness (Khan. et al., 2023). Statistical analysis and simple mathematical models have historically been the fundamental components of KPI forecasting. Time-series analysis and linear regression were popular techniques in previous business contexts (Tadayonrad and Ndiaye, 2023; Le et al., 2018). Despite being successful in linear and stable contexts, these approaches frequently fail in complex and dynamic commercial environments. Machine learning (ML) approaches have been gradually used as a result of the shortcomings of previous methodologies. By using algorithms like decision trees, support vector machines, and simple neural networks, machine learning provided more advanced, data-driven methods for KPI forecasting (Le et al., 2018; El Mazgualdi et al., 2021). More complex and adaptive models were made possible by this shift, which was a major turning point in the development of KPI forecasting.

Dealing with data complexity is one of the main challenges in KPI forecasting (Gurtner and Cook, 2017; Armaki and Mohammed, 2023). In addition to this, the need for real-time KPI analysis has fueled the advancement of forecasting techniques. Research has emphasized the need for models that can quickly adjust to evolving data streams to give decision-makers timely information (Gupta et al., 2016; Zhang et al., 2020; Svensson et al., 2015). Also making accurate forecasts is still a significant challenge (Pietukhov et al., 2023).
The forecasting paradigm shifted with the emergence of machine learning. According to (Bishop, 2006), its preliminary use in KPI forecasting showed encouraging outcomes, especially when dealing with more intricate and non-linear data patterns. Deep learning emerged as a result of these techniques, which opened the door for more advanced methods (LeCun et al., 2015). Sun et al. (Sun and Ge, 2020), explores the utilization of deep learning methods in industrial operations to monitor and forecast essential performance metrics (KPIs). The author suggests a technique known as an ensemble semi-supervised gated stacked autoencoder (ES2GSAE). It is also mentioned that ensemble deep learning, semi-supervised learning, and gated stacked autoencoders together make it easier to guess KPIs in business processes. Some of the deep learning models are briefly explained below.
Linear Network: This model is a straightforward tool suitable for regression tasks, specifically for predicting KPIs. Mathematically, it can be expressed as given in Equation 1.

$$y = Wx + b \tag{1}$$

The output in this case is represented by y, the input vector by x, the weight matrix by W, and the bias vector by b. The model can modify the output independently of the input thanks to the bias matrix b, while the weight matrix W specifies how each input contributes to the output.
Long Short-Term Memory (LSTM) can be effective in predicting KPIs in many fields such as finance, healthcare, and manufacturing (Goodfellow et al., 2016).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

It determines the specific data to exclude from the cell state. The process involves taking the previous hidden state h_{t-1} and the current input x_t, combining them, and then applying a linear transformation using the weight matrix W_f and bias b_f. The sigmoid function, denoted by σ, compresses the output values to a range of 0 to 1. Values approaching 1 indicate that the information should be retained, whereas values approaching 0 indicate that the information should be discarded. Mathematically, this can be expressed as given in Equations 2-7.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{4}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$

$$h_t = o_t * \tanh(C_t) \tag{7}$$

LSTM units process sequential data with several critical components. The forget gate uses the previous concealed state and current input to decide which information to keep and which to delete from the cell state, then applies a sigmoid function. The input gate determines what new cell state information to add.

Recurrent Neural Networks (RNNs) have been employed to forecast Key Performance Indicators (KPIs) in diverse fields, such as speech recognition, natural language processing, and finance (Goodfellow et al., 2016). An RNN can be mathematically expressed as given in Equation 8.

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + b) \tag{8}$$

In this equation, the hidden state at time t, or the RNN's memory component storing data from earlier inputs, is represented by the symbol h_t. The weight matrix that connects the current state, W_{hh}, to the prior hidden state, h_{t-1}, is W_{hh}. The weight matrix for the current input, x_t, is W_{xh}, and the bias is b. The hyperbolic tangent function, or *tanh* , adds non-linearity to the input, enabling the network to recognize intricate patterns.

CNN applied for forecasting Key Performance Indicators (KPIs) in fields where the input data exhibits a spatial arrangement, such as sensor data collected from a manufacturing facility (Brownlee, 2018). A CNN can be mathematically expressed as given in Equation 9.

$$Z = f(W * X + b) \tag{9}$$

Here, Z stands for the output feature map, W for the convolutional filter's (kernel's) weights, X for the input data, b for the bias, and the convolution process. Usually, the function f is a non-linear activation function, such as the Rectified Linear Unit (ReLU).

LSTM + CNN model is a hybrid approach that leverages the advantages of both LSTMs and CNNs. This model can be mathematically expressed as given in Equation 10.

$$C = \mathrm{CNN}(X)\ H_t = \mathrm{LSTM}(C, H_{t-1}) \tag{10}$$

At each time point, CNN may extract features related to product quality or equipment performance; the LSTM then learns how these qualities change over time to help anticipate future maintenance needs or performance.

Transformer models perform better in terms of accuracy, efficiency, and scalability than conventional forecasting techniques. Large datasets may be processed and learned by them, which enables a more sophisticated comprehension of data patterns and more precise forecasting. Furthermore, transformer models' scalability allows them to accommodate growing data volumes, something that older approaches frequently cannot (Brownlee, 2018). It performs especially well in scenarios with high dimensionality, complicated data linkages, or quickly evolving trends. Transformer dominance in these situations is their sophisticated data processing powers and their ongoing capacity to absorb and adjust to new information.

3 METHODOLOGY

The transformer model works on the principle of managing sequential data without the requirement for recurrence or convolution. It was initially created for natural language processing applications. These characteristics make it ideal for predicting KPIs, as KPIs frequently depend on time-series data, which is sequential by nature. The model is a great tool for projecting future KPI values based on historical performance, as it can evaluate complete sequences at once, enabling it to gradually identify complicated patterns. The self-attention mechanism is the main predictive technique used by the transformer. This makes it possible for the model to assign varying weights to various input sequence points. This makes it possible for the model to assign varying weights to various input sequence points. This implies that when predicting future KPI values, the model can determine which historical data points (e.g., past sales numbers, website traffic) are most significant. The model gains knowledge of the correlations and patterns present in the data by training on prior KPI data, which enables it to forecast future performance with accuracy. Transformers analyze all data points concurrently, which greatly reduces training and inference times as compared to recurrent neural networks. It can readily capture correlations between data points that are widely apart in the sequence because of the self-attention process. This is important since long-term trends and patterns are important for KPI prediction. Beyond text, transformers may be used with a variety of data formats, including numerical time-series data which is important for KPI forecasting.

Step 1: Representation of Query, Key, and Value: Three representations are produced given a series of input data KPI values: Values (V), Keys (K), and Queries (Q). These are acquired by input data alterations that are learned linearly. These representations, when used in the context of KPI prediction, encapsulate many facets of the input data that are essential for comprehending its relevance and context.

Step 2: Attention Score Calculation: The attention scores are determined by multiplying the Queries and Keys using a dot product, as given in Equation 11.

$$Scores = QK^T \tag{11}$$

The similarity between each query and key is measured in this step, which helps determine how much weight to give each data point in a prediction.
Step 3: Scaling: To keep the gradients of the softmax from being too tiny, the scores are reduced by the square root of the keys' dimension ($\sqrt{d_k}$) as given in equation 12:

$$Scaled\ Scores = \frac{Scores}{\sqrt{d_k}} \tag{12}$$

Step Four: Softmax: Subsequently, a softmax function is applied to the scaled scores to convert them into probabilities, as specified in Equation 13

$$AttentionWeights = Softmax\left(Scaled\ Scores\right) \tag{13}$$

Using the computed scores as a basis, this phase emphasizes the significance of each value in the sequence.
Step 5: Equivalent Sum: Ultimately, the output of the attention mechanism is generated by computing a weighted sum of the Values, utilizing the attention weights as detailed in Equation 14.

$$Output \ = \ Attention\ Weights\ .\ V \tag{14}$$

The results provide an enhanced representation of the input data by emphasizing the most significant segments for predicting future KPI values. The comprehensive formulation of the Transformer model, as detailed in Equation 15, demonstrates this process.

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{15}$$

The methodology is explained step by step below and is also graphically viewed in Figure 1.

Detail Architecture of the Transformer Model: We develop a transformer-based model given the following scenario: 4 batch sizes, 9 sequences (or feature set) lengths, 768 embedding sizes, and 4 attention heads.
1. Input Sequence: A batch of four sequences, each with nine characteristics, in the shape (4, 9), are the inputs that the model takes.
2. Embedding Layer: The sequence's features are converted into 768-dimensional vectors, giving rise to the shape (4, 9, 768).
3. Positional Encoding: Positional encodings are introduced to preserve sequence order information, maintaining the shape at (4, 9, 768).
4. Transformer Blocks: The model can concentrate on various segments of the input sequence by using four heads for self-attention. These building components allow for sophisticated interactions while preserving the geometry of the input. The output shape of each transformer block is (4, 9, 768).
5. Pooling/Aggregation Layer: This optional stage produces a fixed-size output for each sequence by reducing the sequence dimension. Average pooling is a popular method that yields a shape of (4, 768).
6. Fully Connected Layers: After that, the model goes through fully connected layers processing the pooled output, adjusting the final layer for the regression job. The shape that comes before the output layer is (4, 256) if the last hidden layer has a size of 256.
7. Output Layer: Each sequence's regression output is generated by a single linear neuron, giving rise to an output shape of (4, 1).

4 EXPERIMENTS

The study's experimental evaluation is set up to assess how well different deep learning models predict Key Performance Indicators (KPIs) in different datasets. For each dataset, data were divided into training and test set using an 80/20 ratio Results were evaluated using the following four evaluation measures: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2. Three different datasets were used in this work, each of which offers a different perspective on retail operations and consumer behavior. The datasets contain transactional data from two distinct grocery chains, as well as sales data from the European pharmacy chain Rossmann. When taken as a whole, they include a broad range of characteristics, from daily sales and promotional activities to consumer demographics and purchasing habits, providing a thorough understanding of the elements impacting important performance metrics in the retail industry, The datasets are available at [1], [2], [3].

Overview of Rossmann Store Sales: Sales information from Europe's largest chain of pharmacies, Rossmann, is included in this dataset. This dataset contains daily sales data for 1,115 shops. It is commonly used in research and forecasting competitions to anticipate sales, a crucial KPI for retail operations, and is publicly available. Important characteristics are enumerated below:
Sales: The target variable is a direct reflection of daily sales data and shop performance.

[1] https://www.kaggle.com/competitions/rossmann-store-sales/data

[2] https://www.kaggle.com/datasets/anandku79/kpidashboard

[3] https://www.kaggle.com/code/tatianakushniruk/superstore-sales-profit-analysis/input

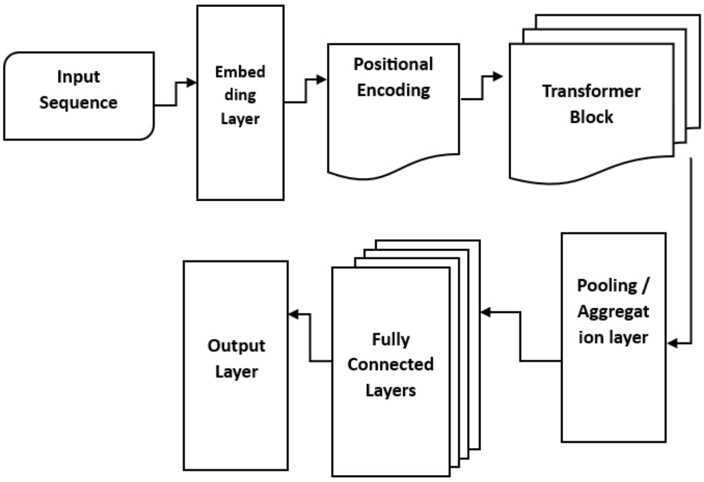


Figure 1: The architecture of the Transformer model.

Store type: Sales trends are influenced by categorical data that indicates the kind of store.
Assortment: categorical information indicating the degree of variety (basic, additional, and extended) that the shop offers and how it affects client decisions and sales volume.
Promotions: Sales numbers are directly impacted by binary data that indicates if a retailer was running a deal on a certain day.
State Holidays: Details on public holidays that have an impact on sales and foot traffic in stores.
Holidays: Binary data indicating the dates of the school holidays may cause fluctuations in sales as a result of shifts in the number of customers.
Day of the Week: An approach to capturing weekly sales patterns using categorical data. Significance to Forecasting KPIs: This dataset makes it possible to investigate the relationship between sales KPIs and a variety of factors, such as holidays, promotions, and shop attributes. The study intends to show how deep learning models can anticipate complicated patterns and temporal correlations present in retail sales data using sales forecasting.
Supermarket Dataset 1: Transactional data from a chain of supermarkets are included in this dataset, which records specific client purchase information over time. Data on product purchases, customer demographics, transaction durations, and modes of payment are all included. The dataset has been organized in a way that makes it easier to analyze consumer behavior and buying trends. Important characteristics are:
Transaction Amount: An important KPI that serves as a direct indicator of income is the monetary value of each transaction.
Product Categories: Detailed data on the categories of goods bought, reflecting consumer inclinations and impacting inventory control tactics.
Customer demographics: Information about a customer's age, gender, and maybe geography that provides hints about potential market niches.
Time of Transaction: Each transaction is timestamped, allowing for the study of peak shopping hours and trends.
Significance for KPI Forecasting: This dataset is used in the study to investigate how well deep learning models predict revenue and comprehend consumer purchase patterns, two essential KPIs for supermarket operations. Future sales patterns and consumer behavior can be predicted by analyzing transaction quantities and customer profiles.
Supermarket Dataset 2: As with Supermarket Dataset 1, this dataset provides a chance to assess the model's performance in other retail contexts by containing transactional and customer data from an alternative supermarket chain or location. While some of the variables may differ in terms of magnitude, client base, or product variety, Supermarket Dataset 1's variables are included. Important characteristics are:
Product Sales Volume: Quantitative information on the quantity of goods sold; essential for supply chain and inventory management.
Promotional Data: Details on in-store sales events, discounts, and exclusive deals that have a big influence on consumer traffic and sales figures.
Customer Loyalty Information: Information on a customer's involvement in loyalty programs that influences the frequency and patterns of their purchases.
Seasonal Variations: Information that captures events and trends specific to a given season is essential for modifying stock and marketing plans.
Relevance to KPI Forecasting: This dataset offers a

Table 1: Experimental results of transformer and other methods over Rossmann Sales.

Algorithms	MAE	MSE	RMSE	R Squared
MLP	835.54	1 750 258.53	1322.57	0.89
LSTM	800.01	1 646 816.85	1283.77	0.91
CNN	1255.74	3 734 871.20	1932.60	0.74
RNN	821.29	1 945 628.38	1394.37	0.88
Transformer	760.48	1 475 395.99	1214.00	0.96

Table 2: Experimental results of transformer and other methods over Supermarket Dataset 1.

Algorithms	MAE	MSE	RMSE	R Squared
MLP	420.64	2 067 452.70	1437.47	0.688
LSTM	347.81	1 898 975.52	1378.01	0.841
CNN	800.00	2 292 541.22	1514.83	0.402
RNN	400.59	1 910 223.67	1382.21	0.590
Transformer	292.10	1 665 437.11	1290.65	0.764

Table 3: Experimental results of transformer and other methods over Supermarket Dataset 2.

Algorithms	MAE	MSE	RMSE	R Squared
MLP	610.39	2 521 637.84	1587	0.81
LSTM	589.10	1 736 824.54	1317	0.91
CNN	810.22	2 944 832.88	1716	0.53
RNN	600.71	2 068 537.32	1438	0.72
Transformer	428.86	1 546 783.55	1243	0.89

benchmark for evaluating the accuracy and flexibility of deep learning models in predicting KPIs in various supermarket environments. It emphasizes how well the models can forecast key performance indicators (KPIs) like sales volume, the success of promotions, and the effect of seasonal fluctuations on sales.

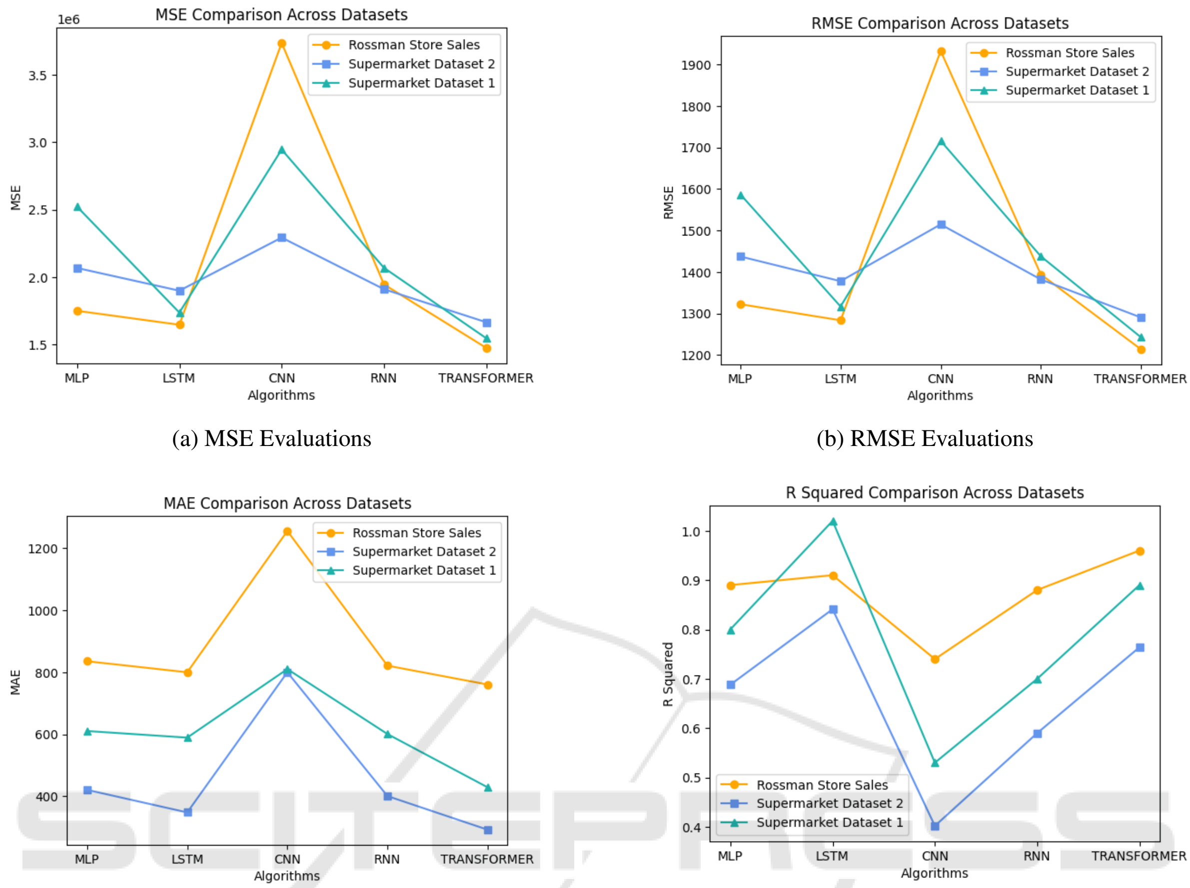
5 RESULTS AND DISCUSSIONS

The results of the above-mentioned experiments show that the Transformer model performs exceptionally well for the Rossmann Store Sales dataset, with the lowest scores for MAE (760.48), MSE (1475395.99), and RMSE (1214.00), as well as the highest R^2 (0.96). All the results are reported in Table 1. Following closely, the LSTM model demonstrates its efficacy in managing sequential data due to its innate ability to retain long-term dependencies. CNNs are normally quite good at extracting features, but they don't seem to work as well in this situation. This could be because sales data is sequential, which means that CNNs aren't built to capture temporal correlations.

Supermarket Dataset 1 provides more evidence of the Transformer model's predictive strength, as it has the lowest MAE (292.10) and RMSE (1290.65) as well as a high R^2 (0.764), indicating its adaptability to a wide range of data types. All the results are reported in Table 2. Due to its ability to handle both sequential and time-series data, the LSTM model performs impressively. Relative to the Rossmann Store Sales dataset, the CNN model performs poorly, potentially for similar reasons. Further highlighting the Transformer and LSTM models' appropriateness for these forecasting tasks is the nature of the data in Supermarket Dataset 1, which is probably rich in temporal patterns like those in Rossmann's dataset.

The Supermarket Dataset 2 study shows a similar pattern, with the Transformer model exhibiting the highest overall performance in terms of all metrics, especially with an impressive R^2 of 0.89. The Transformer model is successful and generalizable in capturing complicated patterns and dependencies in KPI data, as seen by its consistency across datasets. Given the dataset's properties, the LSTM model performs remarkably well, as evidenced by its R^2 of 0.91, which may indicate overfitting or a highly effective model. All the results are reported in Table 3. As seen by their lower performance measures, CNNs once again seem less appropriate for certain forecasting jobs.

Due to the Transformer model's attention mechanism, which successfully captures long-range dependencies and linkages within the data, the findings across all three datasets consistently show the model's high-predicting skills. Not only can LSTM models perform admirably, but their applicability in sequential data processing is also demonstrated. But while the classic CNN and RNN models are well-known for their



(a) MSE Evaluations

(b) RMSE Evaluations

(c) MAE Evaluations

(d) R^2 Evaluations

Figure 2: The evaluation results of different method over various evaluations metrics.

efficacy in other contexts, they seem to be less successful in KPI forecasting. This is probably because KPI data involves complex patterns and unique requirements for recording temporal sequences. These results imply that using advanced deep learning models, like Transformers, for KPI forecasting has a substantial benefit in terms of improved prediction accuracy and a more nuanced comprehension of data patterns. The ability to make good decisions and develop strategic plans based on trustworthy projections has significant ramifications for business intelligence. The comparative study highlights the significance of model selection in forecasting tasks and recommends that the forecasting objectives and data characteristics be carefully considered. Subsequent investigations may examine hybrid models or more advancements in deep learning architectures to further improve forecasting precision and dependability.
For the Supermarket Dataset 1, the MAE of the LSTM (Long Short-Term Memory) model is much higher. This might be a symptom of overfitting or a warning that this specific dataset is not a good fit for this model type, see Figure 2 for more detail. Better model accuracy is implied by lower MSE values, which are desirable. Because the MSE is squared in this case, magnifying differences, the scale is significantly greater. On Supermarket Dataset 1, the MLP and LSTM perform better, comparable to RMSE, but the RNN performs rather poorly, the detail is reported in Figure 2. With 1 denoting perfect prediction, higher R^2 values are preferable. With R^2 values near 1, the MLP and Transformer models obtained the best results, particularly when applied to the Rossmann Store Sales dataset. On the Supermarket Dataset 1, the CNN and RNN exhibit lower R^2 values, suggesting that they may not be adequately capturing the variability of the data.

6 CONCLUSION

In this research work, a variety of deep learning models were used to forecast Key Performance Indicators (KPIs) and compared with the proposed transformer learning model. Forecasting accuracy is significantly higher with transformer models than with other deep learning. One of their greatest advantages is their capacity to handle non-linear interactions, recognize intricate patterns, and analyze vast and varied datasets. This study's comprehensive assessment of deep learning models for KPI forecasting across a variety of datasets highlights the Transformer model's exceptional performance. It achieves an R^2 of 0.96 for prediction accuracy. The accuracy of forecasts is boosted to a new level by this model's capacity to represent detailed temporal interactions. On the other hand, as seen by its lower R^2 values across datasets, the CNN model had shortcomings in processing sequential data. To enhance the model's comprehension of affecting elements, future research may investigate more complex transformer topologies for more accurate KPI predictions. Additionally, other datasets may be considered.

DECLARATIONS

This research has received funding from the project Vitality – Project Code ECS00000041, CUP I33C22001330007 - funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5 - 'Creation and strengthening of innovation ecosystems,' construction of 'territorial leaders in R&D' – Innovation Ecosystems - Project 'Innovation, digitalization and sustainability for the diffused economy in Central Italy – VITALITY' Call for tender No. 3277 of 30/12/2021, and Concession Decree No. 0001057.23-06-2022 of Italian Ministry of University funded by the European Union – NextGenerationEU.

REFERENCES

Armaki, S. and Mohammed, K. A. (2023). Complex kpis versus the usual benchmarks.

Bishop, C. (2006). Pattern recognition and machine learning. *Springer google schola*, 2:531–537.

Brownlee, J. (2018). *Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery.

El Mazgualdi, C., Masrour, T., El Hassani, I., and Khdoudi, A. (2021). Machine learning for kpis prediction: a case study of the overall equipment effectiveness within the automotive industry. *Soft Computing*, 25:2891–2909.

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., et al. (2020). An introductory review of deep learning for prediction models with big data. *Frontiers in Artificial Intelligence*.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Gupta, S., Nikolaou, M., Saputelli, L., and Bravo, C. (2016). Esp health monitoring kpi: a real-time predictive analytics application. In *SPE Intelligent Energy International Conference and Exhibition*, pages SPE–181009. SPE.

Gurtner, G. and Cook, A. (2017). Understanding kpi trade-offs-key challenges of modeling architectures and data acquisition. In *Data Science in Aviation 2017 Workshop*.

Horváthová, J., Mokrišová, M., Suhányiová, A., and Suhányi, L. (2015). Selection of key performance indicators of chosen industry and their application in formation of creditworthy model. *Procedia Economics and Finance*, 34:360–367.

Khan., T., Mircoli., A., Potena., D., and Diamantini., C. (2023). Semantic representation of key performance indicators categories for prioritization. In *Proceedings of the 25th International Conference on Enterprise Information Systems - Volume 2: ICEIS*, pages 142–151. INSTICC, SciTePress.

Le, L.-V., Sinh, D., Tung, L.-P., and Lin, B.-S. P. (2018). A practical model for traffic forecasting based on big data, machine learning, and network kpis. In *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 1–4. IEEE.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

Pietukhov, R., Ahtamad, M., Faraji-Niri, M., and El-Said, T. (2023). A hybrid forecasting model with logistic regression and neural networks for improving key performance indicators in supply chains. *Supply Chain Analytics*, 4:100041.

Sun, Q. and Ge, Z. (2020). Deep learning for industrial kpi prediction: When ensemble learning meets semi-supervised data. *IEEE Transactions on Industrial Informatics*, 17(1):260–269.

Svensson, I., Saeverhagen, E., and Bouillouta, F. M. (2015). Driving rig performance through real-time data analysis, benchmarking, dashboards, and developed key performance indicators. In *SPE digital energy conference and exhibition*, page D021S011R002. SPE.

Tadayonrad, Y. and Ndiaye, A. B. (2023). A new key performance indicator model for demand forecasting in inventory management considering supply chain reliability and seasonality. *Supply Chain Analytics*, 3:100026.

Tsai, Y.-C. and Cheng, Y.-T. (2012). Analyzing key performance indicators (kpis) for e-commerce and internet marketing of elderly products: A review. *Archives of gerontology and geriatrics*, 55(1):126–132.

Zhang, T., Liu, S., Qiu, W., Lin, Z., Zhu, L., Zhao, D., Qian, M., and Yang, L. (2020). Kpi-based real-time situational awareness for power systems with a high proportion of renewable energy sources. *CSEE Journal of Power and Energy Systems*, 8(4):1060–1073.