Enhancing Recurrent Neural Network Efficacy in Online Sales Predictions with Exploratory Data Analysis

Erni Widiastuti[1], Jani Kusanti[2], Herwin Sulistyowati[3]
Faculty of Economic, Universitas Surakarta, Indonesia[1]
Faculty of Electrical Information Engineering, Universitas Surakarta, Indonesia[2]
Faculty of Law, Universitas Surakarta, Indonesia[3]

***Abstract*—Online sales forecasting has become an essential aspect of effective business planning in the digital era. The widespread adoption of digital transformation has enabled companies to collect substantial datasets related to consumer behavior, market trends, and sales drivers. This study attempts to uncover patterns and predict sales growth by utilizing product images and their associated filenames as input. To achieve this, we use EDA combined with LSTM and Gated Recurrent Unit (GRU), which excel in processing sequential data. However, the performance of these networks is significantly affected by the quality of data and the preprocessing methods applied. This study highlights the importance of Exploratory Data Analysis (EDA) and Ensemble Methods in enhancing the efficacy of RNNs for online sales forecasting. EDA plays a crucial role in identifying significant patterns such as trends, seasonality, and autocorrelation while addressing data irregularities such as missing values and outliers. These findings show that integrating EDA substantially improves the performance metrics of RNN, as indicated by the reduction in loss and mean absolute error (MAE) values across training epochs (e.g. loss: 0.0720, MAE: 0.1918 at epoch 15). These results indicate that EDA improves the accuracy, stability, and efficiency of the model, allowing RNN to provide more reliable sales predictions while minimizing the risk of overfitting.**

Keywords*—*Exploratory data analysis; recurrent neural networks; online sales prediction; sequential data; trend patterns

I. Introduction

The digital business era has fundamentally reshaped sales forecasting and management by harnessing the power of big data and advanced analytics [1]. Technology-driven digital transformation has brought significant changes to sales processes and environments, altering their functions and strategies [2][3][4]. Deep learning models, particularly in e-commerce systems, have emerged as essential tools for predicting and enhancing sales outcomes. This trend is emblematic of a broader shift toward digital business, with the e-commerce sector serving as a prime example of this ongoing evolution [4].

Extensive research has compared traditional sales forecasting methods with machine learning techniques, yielding diverse findings. Pustokhina [5] noted that machine learning often surpasses traditional methods in accuracy, yet certain traditional techniques, such as the Holt-Winters method, remain effective under specific conditions. Zhang [6] proposed an innovative model that integrates online reviews and search engine data, significantly enhancing forecasting precision. Furthermore, Bajaj [7] and Cheriyan [8] emphasized the utility of machine learning algorithms in sales forecasting, with Bajaj exploring models such as Linear Regression, K-Neighbors Regressor, XGBoost Regressor, and Random Forest Regressor, while Cheriyan recognized Gradient Boosting as the most effective method for forecasting sales trends.

Deep learning, a focused domain within machine learning, has achieved significant progress in recent years, [9] showcasing its value in solving complex classification challenges. Its capability to derive robust statistical features from data sets it apart as a powerful tool. Research [10] highlighted the critical role of raw data in optimizing machine learning performance. Studies [11][12][13] the evidence presented illustrates that state-of-the-art deep learning methodologies, such as Long-Short Term Memory (LSTM) networks and Convolutional Neural Networks (CNN), substantially exceed the performance of conventional machine learning methodologies in the domain of retail sales forecasting.. Furthermore, evidence from [14][15][16][17][18] reinforces this, demonstrating that deep learning models excel in generating accurate customer predictions for marketing intelligence applications. Together, these findings highlight deep learning's substantial impact on enhancing sales prediction accuracy and operational efficiency.

This study aims to identify sales trends and analyze the factors that influence sales growth within the digital business environment. To fill the existing research gap, we propose an innovative approach that integrates Exploratory Data Analysis (EDA) for feature preprocessing with Ensemble Methods applied to Recurrent Neural Networks (RNNs) to predict product similarity based on e-commerce image data. This research underscores the significant role of sequential learning techniques and vector embedding in enhancing the accuracy of product similarity predictions. Furthermore, EDA is highlighted as a crucial tool for deriving insights and ensuring data quality in the analysis of time series and sequential data.

II. Previous Research Study

Jelonek [19] asserts that Big Data analytics is a vital asset for business management, providing diverse benefits across numerous activities. Ansari [20] highlights how the integration of cloud computing with Big Data analytics ensures cost-effective and scalable solutions for storing and analyzing vast enterprise datasets. Nevertheless, the literature lacks a focused discussion on leveraging deep learning for sales prediction. Alsghaier [21] explains that Big Data analytics equips organizations with actionable insights, enhancing business performance and fostering a competitive edge. Ayuningtyas [22] emphasizes that descriptive, predictive, and prescriptive analytics within Big Data are indispensable for strategic decision-making across industries. Singh [23] investigates the vast possibilities of deep learning within the automotive industry, including its use in self-driving cars, safety systems, virtual sensors, and cutting-edge product development.

Recent developments in research have concentrated on employing deep learning techniques to forecast consumer purchasing behavior. Geetha [15] introduces a deep neural network model that leverages multitask learning to predict consumer preferences by analyzing underlying factors and sentiment. Xia [24] develops a multi-task LSTM model to capture the complexities of the consumer buying decision-making process and estimate purchase probabilities. Nisha [25] conducts a comparative analysis of various neural network architectures, including MLP, LSTM, and TCN, for predicting future purchase behavior in e-commerce. These models consistently outperform traditional machine learning methods across multiple dimensions of consumer behavior prediction.

Liu [26] underscores the transformative impact of computer vision technology on e-commerce platforms, particularly in the realm of sales forecasting. Deep learning models enable the automatic extraction of crucial features from product images, providing valuable insights for predicting sales outcomes. Zhao [27] illustrates how Convolutional Neural Networks (CNNs) can efficiently extract features from structured data, improving forecasting accuracy without requiring manual feature engineering. Qi [28] introduces DSF, a deep neural framework designed to tackle the complexities of promotional activities and product competition in sales forecasting, surpassing traditional baseline models and other deep learning techniques in performance. Yang [29] highlights that the application of computer vision in e-commerce extends far beyond sales forecasting, driving improvements in operational efficiency and customer satisfaction across various areas of online shopping.

Kassem [30] proposed models that classify reviews as either positive or negative and compare them with the ratings provided by users to identify any inconsistencies. These strategies aim to enhance the accuracy and reliability of product information for consumers. Wang [31] developed CLUE, a fraud detection system that leverages recurrent neural networks to analyze user click behavior and detect suspicious transactions.

III. Proposed Research Methodology

The methodology for this study is illustrated in Fig. 1 which depicts the process of creating an Exploratory Data Analysis (EDA) algorithm model utilizing LSTM and GRU to forecast online sales.

In Fig. 1, Flowchart combining Exploratory Data Analysis (EDA) with LSTM and GRU frameworks.

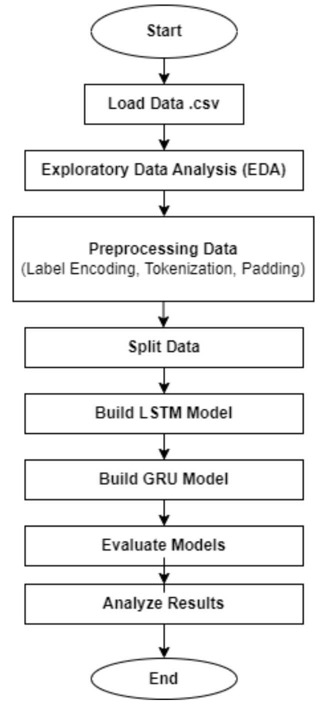

Fig. 1. Flowchart combining exploratory data analysis (EDA) with LSTM and GRU frameworks.

A. Load Data

The initial step involves loading the data from a CSV file. This file, generated during data collection, is stored in a tabular format containing five columns: posting_id, image, image_phash, title, and label_group.

B. Exploratory Data Analysis (EDA)

EDA functions are useful for identifying anomalies, trends, or outliers in sequential data that might obstruct model convergence. It primarily serves as a process for visualizing and understanding data. Below are some commonly used formulas in EDA, for the equation as in (1-6):

- Mean

  The mean reflects the central tendency of a dataset, calculated by summing all individual values and dividing the total by the number of data points.

$$Mean(\mu) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

Where:

x_i = value to- i

$n = the\ total\ amount\ of\ data$

- Variance

Variance quantifies the extent to which data values deviate from the mean.

$$Variance(\sigma^2) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{2}$$

Where:

μ = mean

x_i = value to- i

- Standard Deviation

The standard deviation constitutes the principal square root of the variance.

$$Standar\ Deviation(\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \quad (3)$$

- Covariance

Covariance measures the degree of interdependence between two variables in a dataset.

$$Covariance(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \quad (4)$$

Where:

x_i = value to- i from X

y_i = value to- i from Y

$\overline{x}, \overline{y}$ = mean from X and Y

- Correlation

Pearson correlation quantifies the degree of linear association between two variables.

$$Correlation(\tau) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i-1}^{n}(y_i - \overline{y})^2}} \quad (5)$$

Correlations can vary between −1 and 1, with 1 indicating a perfect positive correlation, 0 denoting no correlation, and −1 representing a perfect negative correlation.

- Autocorrelation assesses the relationship between values at a specific moment and values from a prior moment (lag) for the purpose of correlation.

$$Autocorrelation(k) = \frac{\sum_{i=k+1}^{n}(x_i - \overline{x})(x_{i-k} - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \quad (6)$$

k is the lag, which is how far back in time we look

C. Recurrent Neural Networks (RNN)

Forward Pass

In a standard RNN, the output of a neuron at time t (*h_t*) is influenced by the current input (*x_t*) and the hidden state from the previous time step (*h_t*−1), for the equation as in (7-8).

- State (Hidden State)

$$h_t = \emptyset(W_{xh}.x_t + W_{hh}.h_{t-1} + b_h) \quad (7)$$

Where:

h_t = hidden state at the time t

x_t = input at time t

W_{xh} = weight between input and hidden state

W_{hh} = weight between previous hidden state and the current hidden state

b_h = bias

$\emptyset$ = activation function (usually tanh or ReLU)

- Output

$$y_t = \emptyset(W_{hy}.h_t + b_y) \quad (8)$$

Where:

y_t = output at the time t

W_{hy} = weight between output and hidden state

b_h = bias output

$\emptyset$ = activation function (softmax or sigmoid)

Backpropagation Through Time

The training process employs Backpropagation Through Time (BPTT) to compute the gradient and adjust the weights. The gradient at time step *t* is determined by taking into account all preceding time steps.

The error gradient L for parameter W, for the equation as in (9):

$$\frac{\partial L}{\partial W} = \sum_{t-1}^{T}\frac{\partial L_t}{\partial W} \quad (9)$$

T is the total number of times in the sequence

LSTM is designed to address the vanishing gradient issue commonly encountered in standard RNNs. It employs three primary gates: the input gate, the output gate, and the forget gate, for the equation as in (10-14).

- Forget Gate

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \quad (10)$$

f_t = the forget gate that determines what information will be forgotten

- Input Gate

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (11)$$

i_t = the input gate that determines what new information will be stored

- Update Cell State

$$C_t = f_t.C_{t-1}.i_t \quad (12)$$

C_t = the cell state at time t

- Output Gate

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (13)$$

o_t = the output gate that determines the output of the LSTM at time t

- Hidden State

$$h_t = o_t.\tanh(C_t) \quad (14)$$

h_t = the hidden state or output of the LSTM

IV. Experimental Result and Analysis

During the training of our EDA using both LSTM and GRU models, we began by performing an exploratory analysis on the label groups found in Train.csv, examining their data distributions. The results before using EDA are presented in Fig. 2. While the results after using EDA are presented in Fig. 3.

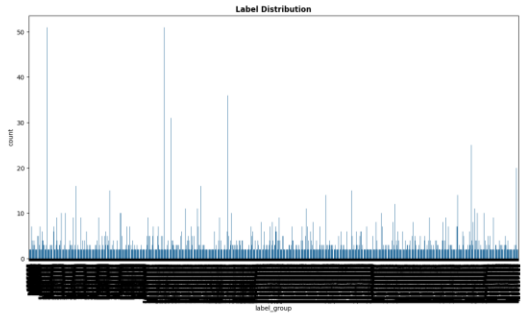

Fig. 2. The distribution of label_group in Train.csv without performing EDA.

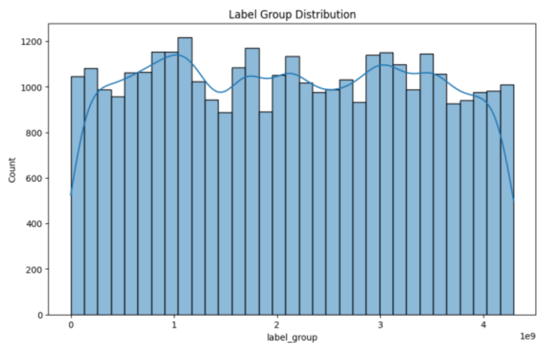

Fig. 3. The distribution of label_group in Train.csv with performing EDA.

From Fig. 2, the figure presents a bar chart. The horizontal axis (x-axis) denotes the label_groups, while the vertical axis (y-axis) displays the count or frequency of occurrences for each label_group. The distribution of label_groups is notably skewed and unbalanced. Some label groups exhibit very high frequencies, surpassing 50, whereas the majority have significantly lower frequencies, frequently below 10, with many falling under 5. The presence of several tall bars indicates that certain label groups are particularly dominant within the dataset. This suggests a bias in the data towards specific labels. Such an imbalanced distribution of labels has important implications for machine learning modeling. If a model were trained directly on this data, it would likely favor the dominant label groups and perform poorly on those that occur less frequently.

From Fig. 3, this figure illustrates the distribution of label_groups in the Train.csv dataset after conducting Exploratory Data Analysis (EDA). Unlike Fig. 2, which shows a significant imbalance, this figure presents a much more balanced distribution. The histogram displays the frequency of each label_group value range as bars, while the KDE (blue line) provides a smoother estimate of the variable's probability distribution. Compared to Fig. 2, the distribution of label_groups here is notably more uniform, with no excessively tall bars. This indicates that the occurrence frequency of each label_group value range is relatively equal. Such a balanced distribution has positive implications for machine learning modeling. The main difference between Fig. 2 and Fig. 3 is the degree of imbalance: Fig. 2 reveals extreme disparity, where some label_groups are highly prevalent while most others are rarely observed. In contrast, Fig. 3 illustrates a more equitable distribution following EDA. This suggests that EDA has resulted in a significant transformation of the data, possibly involving clustering or other modifications to the label_groups.

The test results to explore the distribution of text length in title_image in text.csv can be shown in Fig. 4.

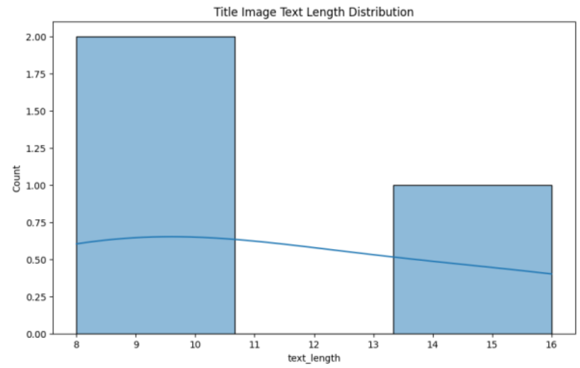

Fig. 4. Title image text length distribution.

Fig. 4 depicts the distribution of lengths for image caption texts. The X-axis represents text lengths, which range from about 8 to 16 characters or words. This shows that the lengths of the image captions fall within this range. The Y-axis indicates the frequency of occurrences for each range of text lengths.

	label_group
count	3.425000e+04
mean	2.128611e+09
std	1.234630e+09
min	2.580470e+05
25%	1.050720e+09
50%	2.120410e+09
75%	3.187910e+09
max	4.294197e+09

Fig. 5. The results of the EDA calculation.

Fig. 5 shows the EDA calculation results. Fig. 6 displays the results of the Exploratory Data Analysis (EDA) for the label_group variable. Below is a detailed explanation of the statistics:

- Count: The value 3.425000e+04 (34,250) indicates that there are 34,250 data points or observations in the label_group variable, representing the total number of entries analyzed.
- Mean: The value 2.128611e+09 (2,128,611,000) represents the average of all label_group values, providing insight into the central tendency of the data.
- Standard Deviation (std): The value of 1.234630e+09 (1,234,630,000) indicates the extent to which the data varies from the mean. A high standard deviation implies that the data points are significantly dispersed, whereas a

low standard deviation suggests that they are closely grouped around the mean. In this instance, the large standard deviation in relation to the mean indicates substantial variability within the data.

- Minimum (min): The value 2.580470e+05 (258,047) denotes the smallest value in the label_group variable.
- First Quartile (25%): The value 1.050720e+09 (1,050,720,000) represents the first quartile. This indicates that 25% of the data falls below or is equal to this value.
- Median (50% / Second Quartile Q2): The value 2.120410e+09 (2,120,410,000) is the median or second quartile. This is the midpoint of the dataset; half of the values are below this point and half are above it. Notably, the median (2.120410e+09) is very close to the mean (2.128611e+09), suggesting a relatively symmetric distribution despite a large standard deviation.
- Third Quartile (75%): The value 3.187910e+09 (3,187,910,000) indicates that 75% of the data has values less than or equal to this figure.
- Maximum (max): The value 4.294197e+09 (4,294,197,000) represents the largest value in the label_group variable.

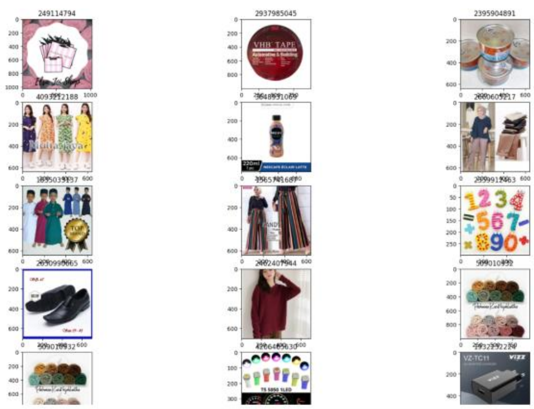

Fig. 6. Example image from label_group.

After the EDA process results, continued with the RNN process which is continued by carrying out the Split Data process for Training and Validation.

TABLE I. RESULTS OF THE SPLIT DATA PROCESS FOR TRAINING AND VALIDATION

Description	Result
Training data shape	(27400, 100) (27400,)
Validation data shape	(6850, 100) (6850,)

Table I shows the results of dividing the training data into Training Data and Validation Data. The results of the Training Data (X_train, y_train) are shown in Table I: (27400, 100) (27400,) (27400, 100). This signifies that the training dataset consists of 27,400 samples or observations, with each sample containing 100 features or time steps (27400,). This represents the shape of the labels or target variables for the training data, indicating that there are 27,400 labels, corresponding to one label for each sample in the training set.

The results of the Validation Data (X_val, y_val) are shown in Table I: TABLE I. (6850, 100) (6850,). (6850, 100). This indicates that the validation dataset consists of 6,850 samples, with each sample containing 100 features or time steps, similar to the training data. The validation data is utilized to assess the model's performance during training and to mitigate the risk of overfitting. (6850,): This represents the shape of the labels or target variables for the validation data, indicating there are 6,850 labels—one corresponding to each validation sample.

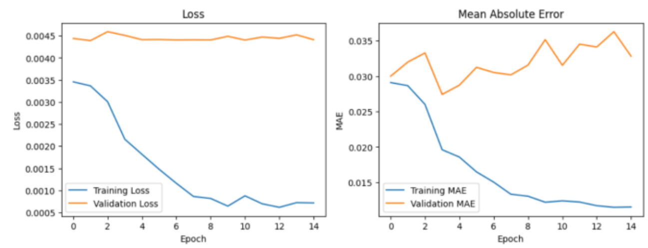

Fig. 7. Training Loss and MAE results with LSTM model without EDA.

In Fig. 7, the document exhibits two graphical representations that monitor the efficacy of the model throughout the training process: the Loss graph and the Mean Absolute Error (MAE) graph, both of which are charted in relation to the epochs.

1) Loss graph: X-axis (Epoch): Represents the number of training epochs, ranging from 0 to 14. Each epoch corresponds to a complete iteration in which the model processes the entire training dataset. Y-axis (Loss): Displays the loss value, which measures the discrepancy between the model's predictions and the actual values. A lower loss value indicates better model performance. Training Loss (Blue Line): Sharp Decline at the Start: The training loss experiences a rapid decrease from approximately 0.0035 at epoch 0 to below 0.001 around epoch 5. This indicates that the model is quickly learning and enhancing its performance on the training data in the early stages. Relatively Stable After Epoch 5: Following epoch 5, while the training loss continues to decline, the rate of decrease becomes less significant and exhibits slight fluctuations. Validation Loss (Orange Line): Generally Stable: The validation loss remains relatively stable at around 0.0045 throughout the training process. Although there are minor fluctuations, there is no clear downward trend as seen in the training loss. A notable gap exists between training loss and validation loss, with the training loss being significantly lower than the validation loss.

2) Mean absolute error (MAE) graph: X-Axis (Epoch): Similar to the loss graph, this axis represents the number of training epochs. Y-Axis (MAE): This axis indicates the MAE value, which measures the average absolute difference between the model's predictions and the actual values. A lower MAE signifies better model performance. Training MAE (Blue Line): Sharp Decline at the Start: Like the training loss, the training MAE also decreases rapidly during the initial stages of training. Relatively Stable After Epoch 5: After epoch 5, the training

MAE stabilizes and shows only minor fluctuations. Validation MAE (Orange Line): Initial Increase Followed by Fluctuations: The validation MAE experiences a slight increase at the beginning of training and then tends to fluctuate around 0.032 to 0.036, without a consistent downward trend. There is a considerable difference between the training MAE and validation MAE, with the training MAE consistently being lower than the validation MAE.

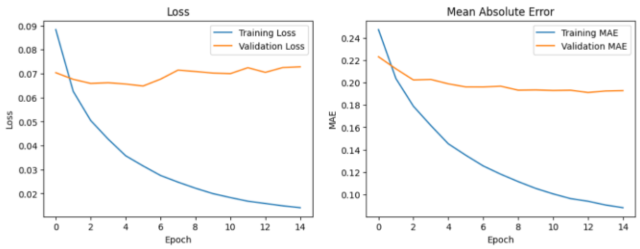

Fig. 8. Training Loss and MAE results with the LSTM model after using EDA.

Fig. 8 presents two graphs that track the model's performance during training, utilizing the Loss and Mean Absolute Error (MAE) metrics plotted against epochs. These metrics are commonly employed to assess regression models. Below are detailed descriptions of each graph.

3) Loss graph: X-axis (Epoch): Displays the number of training epochs, ranging from 0 to 14. Each epoch represents a complete iteration in which the model processes the entire training dataset. Y-axis (Loss): Represents the loss value, which measures the extent to which the model's predictions deviate from the actual values. A lower loss value indicates better model performance. Training Loss (Blue Line): Rapid Decline at the Start: The training loss decreases quickly and consistently from approximately 0.09 at epoch 0 to below 0.02 at epoch 14. This suggests that the model is effectively learning and enhancing its performance on the training data. Consistent Decrease: This steady decline indicates that the model continues to learn throughout the epochs. Validation Loss (Orange Line): Relatively Stable After Initial Decrease: The validation loss shows a slight decrease at the beginning of training, but after a few epochs, it stabilizes and fluctuates slightly around 0.07. There is a notable difference between training loss and validation loss, with training loss consistently lower than validation loss. This difference remains small and stable from the midpoint to the end of the epochs.

4) Mean absolute error (MAE) graph: X-axis (Epoch): Similar to the loss graph, this axis indicates the number of training epochs. Y-axis (MAE): Displays the MAE value, which measures the average absolute difference between the model's predictions and actual values. A lower MAE signifies better model performance. Training MAE (Blue Line): Sharp and Consistent Decline: The training MAE decreases rapidly and steadily from around 0.25 at epoch 0 to below 0.10 at epoch 14. This trend parallels the decrease in training loss, indicating an improvement in model performance on the training data. Validation MAE (Orange Line): Relatively Stable After Initial Decrease: Similar to validation loss, validation MAE experiences a slight initial decrease before remaining relatively stable, fluctuating around 0.20 from mid-epoch to the end. A significant difference exists between training MAE and validation MAE, with training MAE consistently lower. This difference is small and stable from mid-epoch to end.

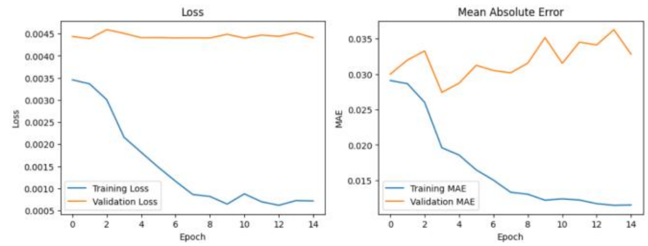

Fig. 9. Training Loss and MAE results with the GRU model after using EDA.

Fig. 9 below presents two graphs that track the model's performance during training, utilizing Loss and Mean Absolute Error (MAE) metrics plotted against epochs. These metrics are commonly employed to assess regression models. Below is a detailed description:

5) Loss graph: X-axis (Epoch): Displays the number of training epochs, ranging from 0 to 14. Each epoch represents a complete iteration during which the model processes the entire training dataset. Y-axis (Loss): Represents the loss value, which quantifies how poorly the model's predictions align with the actual values. A lower loss value indicates better model performance. Training Loss (Blue Line): Sharp Early Decline: The training loss decreases rapidly from approximately 0.0035 at epoch 0 to below 0.001 around epoch 5, indicating that the model is quickly learning and enhancing its performance on the training data in the early stages. Relatively Stable After Epoch 5: After epoch 5, while the training loss continues to decline, the rate of decrease becomes less significant and shows slight fluctuations. Validation Loss (Orange Line): Generally Stable: The validation loss remains relatively stable at around 0.0045 throughout the training process, exhibiting minor fluctuations without a clear downward trend as seen in the training loss. A notable difference exists between training loss and validation loss, with training loss consistently being much lower than validation loss.

6) Mean absolute error (MAE) graph: X-axis (Epoch): Similar to the loss graph, this axis indicates the number of training epochs. Y-axis (MAE): Displays the MAE value, which measures the average absolute difference between the model's predictions and actual values. A lower MAE signifies improved model performance. Training MAE (Blue Line): Sharp Decline at the Start: Like the training loss, the training MAE decreases rapidly during the initial phase of training. Relatively Stable After Epoch 5: Following epoch 5, the training MAE also stabilizes and exhibits only minor fluctuations. Validation MAE (Orange Line): Initial Increase Followed by Fluctuations: The validation MAE shows a slight increase at the beginning of training before fluctuating around

0.042 to 0.046, without a consistent downward trend. There is a significant difference between training MAE and validation MAE, with training MAE consistently lower than validation MAE.

7) The model evaluation results highlight two key metrics: Mean Squared Error (MSE) and Mean Absolute Error (MAE). MSE quantifies the average of the squared differences between the model's predictions and the actual values, with a lower MSE indicating greater accuracy in predicting the target value. On the other hand, MAE assesses the average of the absolute differences between the model's predictions and the actual values, where a lower MAE also signifies improved accuracy in predicting the target value. In this study, three tests were conducted, and the results are presented in Table II.

TABLE II. THE MODEL EVALUATION RESULTS

Epoch Value	Test Results	
	MSE	MAE
LSTM without EDA	2.7837	2.818866014480591
LSTM with EDA	0.07274550944566727	0.1918681412935257
GRU with EDA	0.0053	0.046813491731882095

From Table II, it can be observed that the results for the LSTM model without EDA show a significant disparity between the training metrics (both loss and MAE) and the validation metrics, indicating a strong likelihood of overfitting. The model performs exceptionally well on the training data but struggles with the validation data that it has not encountered before. In contrast, the results for the LSTM model with EDA demonstrate a consistent decline in both training loss and training MAE, suggesting effective learning on the training data. Although there is a difference between the training and validation metrics, this difference is relatively small and stable from the midpoint of the epochs to the end, indicating that overfitting may be minimal or effectively managed. On the other hand, the results for the GRU model with EDA reveal a considerable difference between the training MAE and validation MAE, with training MAE consistently lower than validation MAE. This significant gap between the training metrics (both loss and MAE) and validation metrics strongly suggests overfitting. The model learns very well from the training data but performs poorly on unseen validation data, as evidenced by the continuous decrease in training loss/MAE while validation loss/MAE remains stable or even increases.

V. CONCLUSION

The extensive results of this study indicate that EDA significantly improves the preparation of sequential data for processing by RNNs, allowing essential patterns in the data to be identified and incorporated into the modeling process. By employing EDA techniques specifically for LSTM and GRU models, the risk of overfitting can be reduced. As a result, EDA positively impacts the performance, stability, and interpretability of the RNN model, helping to minimize biases and variances in predictions. The findings—loss: 0.0724; MAE: 0.1913—at epoch 15 show that, in general, higher epoch numbers correspond to lower loss and MAE values. However, it is important to note that if training loss continues to decrease while validation loss begins to rise, this may indicate overfitting. This situation suggests that the model has become too dependent on the specifics of the training data and is failing to develop the robust capabilities needed for accurate predictions on new datasets.

REFERENCES

[1] T. Lips, "Enhanced Sales Management: Using Digital Forecasting," in Performance Management in Retail and the Consumer Goods Industry, 2019, pp. 247–256.

[2] H. Fischer, S. Seidenstricker, and J. Poeppelbuss, "The triggers and consequences of digital sales: a systematic literature review," J. Pers. Sell. Sales Manag., vol. 43, no. 1, pp. 5–23, 2023.

[3] E. Widiastuti, M. Nugroho, and A. Halik, "The Effect of Service Quality, Brand Image and Social Media on Choosing Decisions With E-Wom and Attitude as Intervening Variables on Students of Private …," The Seybold Report. admin369.seyboldreport.org, 2023.

[4] S. M. Amaninder Kaur, "Digital Transformation in Business Era," Manag. Issues Digit. Transform. Glob. Mod. Corp., 2021.

[5] I. V. Pustokhina and D. A. Pustokhin, "A Comparative Analysis of Traditional Forecasting Methods and Machine Learning Techniques for Sales Prediction in E-commerce," Am. J. Bus. Oper. Res., vol. 10, no. 2, pp. 39–51, 2023.

[6] C. Zhang, Y. X. Tian, and Z. P. Fan, "Forecasting sales using online review and search engine data: A method based on PCA–DSFOA–BPNN," Int. J. Forecast., vol. 38, no. 3, pp. 1005–1024, 2022.

[7] K. Saraswathi, N. T. Renukadevi, S. Nandhinidevi, S. Gayathridevi, and P. Naveen, "Sales prediction using machine learning approaches," AIP Conf. Proc., vol. 2387, no. June, pp. 3619–3625, 2021.

[8] S. Cheriyan, S. Ibrahim, S. Mohanan, and S. Treesa, "Intelligent Sales Prediction Using Machine Learning Techniques," Proc. - 2018 Int. Conf. Comput. Electron. Commun. Eng. iCCECE 2018, pp. 53–58, 2018.

[9] C. A. Leke and T. Marwala, "Introduction to Deep Learning BT - Deep Learning and Missing Data in Engineering Systems," C. A. Leke and T. Marwala, Eds. Cham: Springer International Publishing, 2019, pp. 21–40.

[10] R. Hanocka and H. T. D. Liu, "An introduction to deep learning on meshes," ACM SIGGRAPH 2021 Courses, SIGGRAPH 2021, 2021.

[11] P. M. J. Samonte, E. Britanico, K. E. M. Antonio, J. E. J. Dela Vega, T. J. P. Espejo, and D. C. Samonte, "Applying Deep Learning for the Prediction of Retail Store Sales," no. 20, pp. 1–11, 2023.

[12] K. Geetha, "Deep Learning and Sentiment Analysis Improve E-commerce Sales Prediction," 2023 International Conference on Data Science and Network Security, ICDSNS 2023. 2023.

[13] X. Yin, "Prediction of Merchandise Sales on E-Commerce Platforms Based on Data Mining and Deep Learning," Sci. Program., vol. 2021, 2021.

[14] D. Wang, H. Xiao, Q. Sun, and Y. Chen, "The application of deep learning algorithm in marketing intelligence," Proc. 2019 IEEE 3rd Adv. Inf. Manag. Commun. Electron. Autom. Control Conf. IMCEC 2019, no. Imcec, pp. 1356–1360, 2019.

[15] M. P. Geetha, "Deep learning architecture towards consumer buying behaviour prediction using multitask learning paradigm," J. Intell. Fuzzy Syst., vol. 46, no. 1, pp. 1341–1357, 2024.

[16] H. Zhu, "A deep learning based hybrid model for sales prediction of E-commerce with sentiment analysis," Proceedings - 2021 2nd International Conference on Computing and Data Science, CDS 2021. pp. 493–497, 2021.

[17] S. Wang, "M-GAN-XGBOOST model for sales prediction and precision marketing strategy making of each product in online stores," Data Technol. Appl., vol. 55, no. 5, pp. 749–770, 2021.

[18] X. Ma, "E-Commerce Review Sentiment Analysis and Purchase Intention Prediction Based on Deep Learning Technology," J. Organ. End User Comput., vol. 36, no. 1, pp. 1–29, 2024.

[19] D. Jelonek, "Big Data Analytics in the Management of Business," MATEC Web Conf., vol. 125, pp. 1–6, 2017.

[20] S. M. Ansari, "A Review Paper On Big Data Analytics in Cloud," Int. J. Adv. Res. Sci. Commun. Technol., vol. 4, no. 11, pp. 477–483, 2021.

[21] H. Alsghaier, "The Importance of Big Data Analytics in Business: A Case Study," Am. J. Softw. Eng. Appl., vol. 6, no. 4, p. 111, 2017.

[22] A. Ayuningtyas, S. Mokodenseho, A. M. Aziz, D. Nugraheny, and N. D. Retnowati, "Big Data Analysis and Its Utilization for Business Decision-Making," West Sci. Inf. Syst. Technol., vol. 1, no. 01, pp. 10–18, 2023.

[23] K. B. Singh and M. A. Arat, "Deep Learning in the Automotive Industry: Recent Advances and Application Examples," pp. 1–14, 2019.

[24] Q. Xia, P. Jiang, F. Sun, Y. Zhang, X. Wang, and Z. Sui, "Modeling consumer buying decision for recommendation based on multi-task deep learning," Int. Conf. Inf. Knowl. Manag. Proc., pp. 1703–1706, 2018.

[25] Nisha, "Customer Behavior Prediction using Deep Learning Techniques for Online Purchasing," 2023 2nd International Conference for Innovation in Technology, INOCON 2023. 2023.

[26] W. D. Liu, "Application of Computer Vision on E-Commerce Platforms and Its Impact on Sales Forecasting," J. Organ. End User Comput., vol. 36, no. 1, 2024.

[27] K. Zhao and C. Wang, "Sales Forecast in E-commerce using Convolutional Neural Network," Retrieved from http//arxiv.org/abs/1708.07946, 2017.

[28] Y. Qi, C. Li, H. Deng, M. Cai, Y. Qi, and Y. Deng, "A Deep Neural Framework for Sales Forecasting in E-Commerce," pp. 299–308, 2019.

[29] C. Yang and Z. Liu, "Application of Computer Vision In Electronic Commerce," 2021.

[30] M. A. Kassem, "A novel deep learning model for detection of inconsistency in e-commerce websites," Neural Comput. Appl., 2024.

[31] S. Wang, C. Liu, X. Gao, H. Qu, and W. Xu, "Session-Based Fraud Detection in Online E-Commerce Transactions Using Recurrent Neural Networks," Springer Nat. Link, vol. 10536, no. Computer Science, pp. 241–252, 2017.