Predictive Analysis about E-Commerce Text Messages

Jose Ignacio Armas
Facultad de Ingeniería
Universidad del Pacífico
Lima, Perú
ji.armasf@alum.up.edu.pe

César Cabezas
Facultad de Ingeniería
Universidad del Pacífico
Lima, Perú
cm.cabezasg@alum.up.edu.pe

Miguel Muñoz
Facultad de Ingeniería
Universidad del Pacífico
Lima, Perú
ma.muñozc@alum.up.edu.pe

***Abstract*—The rapid digitization driven by the COVID-19 pandemic has significantly transformed the retail sector, with e-commerce witnessing unprecedented growth. As businesses pivot to online sales channels, understanding and optimizing sales interactions has become critical. This study explores the application of Natural Language Processing (NLP) techniques to analyze and predict outcomes in sales conversations. Leveraging a public dataset of sales interactions labeled by success outcomes, we developed a predictive model using Long Short-Term Memory (LSTM) networks and Word2Vec embeddings. Our findings demonstrate that deep learning models significantly outperform traditional machine learning approaches in predicting sales outcomes, with LSTMs achieving the highest accuracy and F1 scores. These results highlight the potential of NLP-driven strategies to enhance sales efficiency and customer relationship management. Future research should focus on expanding datasets, incorporating multi-lingual capabilities, and integrating real-time analysis for broader applicability.**

***Index Terms*—natural language processing, sales prediction, deep learning, LSTM, Word2Vec, e-commerce, machine learning**

I. Introduction

Massive digitization driven largely by the COVID-19 pandemic has impacted different sectors significantly, the retail sector being one of them. With the permanent or temporary closure of physical stores and mobility restrictions, both businesses and consumers have had to migrate to e-commerce and virtual interactions [1]. According to an UNCTAD study in 2021, the global volume of e-commerce increased by more than 20% during the first year of the pandemic, reaching USD 26.7 trillion [2]. In the case of Peru, before the pandemic, online purchases reached USD 6 billion, and in 2021 reached USD 9.3 billion, an annual growth of 55% over the previous year [3]. This increase reflects the global reliance on digital platforms for transactions, forcing companies to optimize their online sales channels.

Technologies such as mobile applications and a variety of digital channels for communication have allowed consumers to have a wide range of alternatives for the purchase of goods or services [4]. One of the most important challenges in this new context is the ability to efficiently analyze and understand the interactions carried out on these platforms. Due to the nature of their work, sales consultants must serve multiple clients at once to increase their sales figures, which can affect their performance [5]. Sales advisors on average make 12 transactions simultaneously, and can reach a maximum of 28 at the same time. To improve efficiency without the need to hire more staff, the use of techniques such as written or sound pattern recognition, Speech Analytics, has allowed advisors to focus on the most promising cases [6].

In this context, natural language processing emerges as a field that investigates how computers can understand, interpret, and manipulate human language in both text and speech formats [7]. It bridges the gap between human communication and machine understanding, enabling systems to process vast amounts of unstructured language data efficiently which required some sort of text preprocessing. It often includes steps such as removing special characters, punctuation and other non-textual elements to enhance the performance of models which use this data [8].

This is the case in the analysis of sales conversations. We propose to develop a Machine Learning model that allows the detection of patterns and the evaluation of the quality of the conversations (intervals according to the predictive propensity) during the conversation; thus supporting the consultant in his work, as well as highlighting the impact of his answers on the sales prospect. In this sense, we use a public database of message-by-message sales conversations, labeled as to whether they ended in sales or not. The implementation of this model would allow companies to know how to direct sales conversations, identifying patterns that correlate positive results and allowing them to optimize strategies.

II. Literature Review

Recent advancements in Natural Language Processing (NLP) have facilitated the automatic analysis of textual interactions between customers and businesses. Specifically, the ability to predict sales outcomes based on customer-seller conversations has gained significant traction in the field of customer relationship management and e-commerce. The use of conversational data allows companies to make informed decisions about customer behavior and potential sales, aligning with the broader trend of using AI and machine learning to automate business processes [9].

Research in this domain often employs Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and

Gated Recurrent Unit (GRU) models due to their capacity to handle sequential data, such as conversations. Bakker [10] explored the use of LSTM and GRU algorithms to predict shopping intentions from clickstream data, highlighting the effectiveness of these models in e-commerce settings where sequential user interactions are critical to predicting outcomes [10]. In these studies, the LSTM model often outperforms other algorithms due to its ability to mitigate the vanishing gradient problem and retain long-term dependencies in sequential data [16].

In addition to analyzing sequences, sentiment analysis has emerged as a pivotal technique for interpreting customer sentiments and correlating them with sales outcomes. By assessing the polarity of customer comments, businesses can gauge the likelihood of completing a transaction. Recent work by Alshamari [12] demonstrated that sentiment analysis, when combined with deep learning models, can significantly improve the accuracy of sales predictions in customer-centric industries [12]. This is particularly useful in detecting the satisfaction or dissatisfaction of a customer during an interaction, which often precedes a purchasing decision.

Sentiment analysis is a vital application in e-commerce, where customer reviews can heavily influence purchasing decisions. Singh et al. [13] explored various machine learning and deep learning models, including BERT, ELMo, and FastText, to predict the sentiment of e-commerce website reviews using an Amazon product review dataset [13]. Their analysis demonstrates that the FastText model combined with a multi-channel Convolutional Neural Network (CNN) achieved the highest validation accuracy of 79.83%. This work highlights the importance of automated sentiment analysis in understanding customer feedback and improving service offerings, particularly in online marketplaces where large volumes of data need to be processed efficiently.

Automated customer service systems, such as chatbots, have become integral to predicting and responding to customer inquiries. These systems utilize NLP models to analyze queries and provide tailored responses, thereby enhancing the customer experience and influencing purchasing decisions [9]. Studies have shown that automating these interactions not only increases efficiency but also allows businesses to detect potential sales opportunities earlier in the customer journey [14]. Furthermore, using Named Entity Recognition (NER) and Part-of-Speech (POS) tagging, NLP systems can identify key components in conversations, such as product names or intent markers, which play a crucial role in predicting the outcome of a sale.

III. Theoretical Framework

Natural Language Processing serves as the foundation for automating customer interactions by enabling machines to understand and generate human language. **Text mining** is a crucial aspect of NLP, involving the extraction of meaningful information from unstructured text data [12]. This process allows businesses to analyze customer inquiries, detect sentiments, and identify key topics, which is essential for providing personalized responses and enhancing customer satisfaction.

Word embeddings, such as those generated by **Word2Vec** models, represent words in continuous vector spaces where semantic relationships are preserved [15]. Introduced by Mikolov et al., Word2Vec uses neural networks to learn word associations from large datasets, capturing context and meaning in a way that traditional bag-of-words models cannot. This technique is instrumental in understanding the nuances of customer language and intent.

Recurrent Neural Networks (RNNs) are designed for processing sequential data, making them suitable for language modeling tasks. However, standard RNNs struggle with long-term dependencies due to the vanishing gradient problem [16]. To address this, **Long Short-Term Memory (LSTM)** networks were developed. LSTMs introduce memory cells and gating mechanisms that regulate the flow of information, allowing the network to retain information over extended time periods.

The LSTM architecture involves several components:

$$
\begin{aligned}
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad \text{(Input Gate)}\\
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad \text{(Forget Gate)}\\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad \text{(Output Gate)}\\
\tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad \text{(Cell Input)}\\
c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad \text{(Cell State Update)}\\
h_t &= o_t \odot \tanh(c_t) \quad \text{(Hidden State)}
\end{aligned}
$$

In these equations:

- x_t is the input vector at time t,
- h_{t-1} is the previous hidden state,
- c_{t-1} is the previous cell state,
- i_t, f_t, and o_t are the input, forget, and output gates, respectively,
- $\tilde{c}_t$ is the candidate cell state,
- W and U are weight matrices,
- b is the bias vector,
- σ denotes the sigmoid function,
- $\tanh$ is the hyperbolic tangent function,
- $\odot$ represents element-wise multiplication.

Gated Recurrent Units (GRUs), simplify the LSTM architecture by combining the input and forget gates into a single update gate [17]. This reduction in complexity often leads to faster training times while maintaining similar performance levels, making GRUs a popular choice for modeling sequential data in NLP tasks.

Apart from neural network approaches, traditional machine learning algorithms like **Decision Trees** and **Support Vector Classifiers (SVC)** are also employed in NLP. Decision Trees are appreciated for their interpretability and ease of use, making them suitable for classification and regression tasks [18]. They work by recursively partitioning the data space and fitting simple prediction models within each partition.

Support Vector Classifiers, based on the work of Cortes and Vapnik, are effective for classification tasks in high-dimensional spaces [19]. SVCs aim to find the hyperplane that best separates data points of different classes by maximizing the margin between them. In the context of text classification, SVCs handle sparse and high-dimensional feature spaces efficiently, which is common when working with textual data represented through techniques like bag-of-words or term frequency-inverse document frequency (TF-IDF).

By integrating these advanced methodologies—ranging from word embeddings and deep learning architectures to traditional machine learning algorithms—businesses can develop sophisticated systems capable of understanding and predicting customer behavior. This comprehensive theoretical framework supports the creation of automated customer service solutions that not only respond to inquiries but also proactively engage customers, ultimately influencing purchasing decisions and enhancing the overall customer experience [14].

IV. Methodology

Figure 1 illustrates the methodology adopted in this study. The dataset used in our research was acquired from a contact center, but due to privacy concerns, the name of the center cannot be disclosed. The dataset was retrieved through queries made directly to the datacenter.

Subsequently, a preprocessing step was carried out on the textual data, which consists of messages in Spanish. After cleaning the data, an exploratory analysis was conducted to uncover patterns and insights within the text. Next, vectorization techniques were applied to convert the text data into numerical formats suitable for machine learning algorithms. The dataset was then divided into two subsets: a training set and a test set. The model was trained using the training data, and its performance was evaluated using the test data.

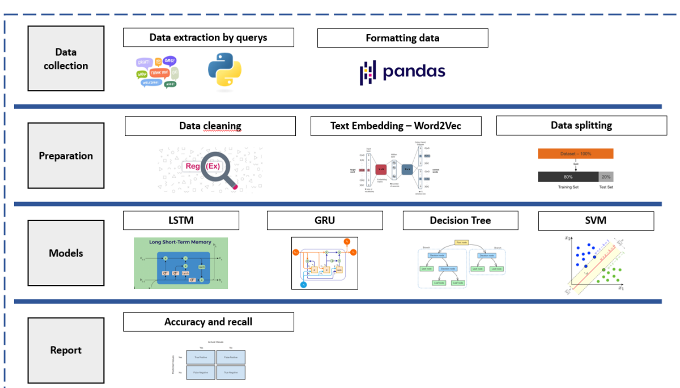

Fig. 1. Methodology of the study

A. Data acquisition

The dataset comes from a POS company or payment gateway, specifically from a third-party contact center that specializes in selling physical equipment. The information from each individual message is stored on a specialized platform, which offers an API for downloading the data, with safeguards such as safe-fail mechanisms. Additionally, it is possible to access the platform's web page using admin credentials. The dataset spans a period of nine months, from January to September, being a total of 4.859 million individual messages.

The dataset includes three main inputs we will use:

- The text of the conversation.
- The speaker, which can be identified as 0 (advisor), 1 (client), or 2 (bot).
- The time difference between consecutive messages.

B. Data preprocessing

In the preprocessing stage, the main goal was to clean and standardize the text data. This involved the following steps:

- Conversion of all text to lowercase to ensure uniformity.
- Removal of punctuation, accents, and special characters, which could introduce noise in the analysis.
- Tokenization of the text into individual words for further processing.

These steps allowed us to prepare the text data for subsequent analysis and modeling by reducing inconsistencies and irrelevant elements.

The data will also require to be joined for each individual string of conversation. That would require use the usage of the phone number as key, but also to separate different sales by defining a new conversation whenever the previous one had a ”final sale message” thanking the user for their purchase. In adition to that, it would be ideal to remove conversations that are deemed too short to give any insights to the model, therefore a filter of at least 10 messages would be put in place.

Those steps would produce a dataset of **143k unique conversations** is produced. In order to mantain similar dimensions in the dataset, we will only consider the later 200 messages for each conversation. This number is chosen due to it being twice the average but half the max length of the conversations that ended in a purchase.

C. Data exploration

Before diving into modeling, an exploratory data analysis (EDA) was conducted to understand the structure and characteristics of the dataset. Key visualizations used include:

- **Class Distribution**: A graphical representation of the different classes in the dataset to assess whether there was any imbalance.
- **Word Clouds**: These were generated to visualize the most frequent terms in the messages, providing insights into the common topics discussed.

The exploration phase provided a clear picture of the dataset, enabling us to adjust our preprocessing and modeling strategies accordingly.

D. Word embedding

To convert the textual data into a format suitable for machine learning models, we used word embedding techniques. Specifically, we applied the Word2Vec algorithm to create dense vector representations of the words. The key configurations explored were:

- **Vector Size**: The dimensionality of the word vectors, which impacts the richness of the word representations.
- **Window Size**: The number of words around a given word considered for context during training.

This process allowed us to represent each word as a vector of real numbers, capturing its semantic meaning based on its context in the dataset.

E. Data splitting

Once the text was vectorized, the dataset was split into two parts:

- **Training Set**: 80% of the data was used to train the machine learning models.
- **Test Set**: The remaining 20% was reserved for evaluating the model's performance on unseen data.

This split ensures that the model is evaluated in a realistic setting, preventing overfitting and providing a measure of its generalization ability.

F. Model and Evaluation

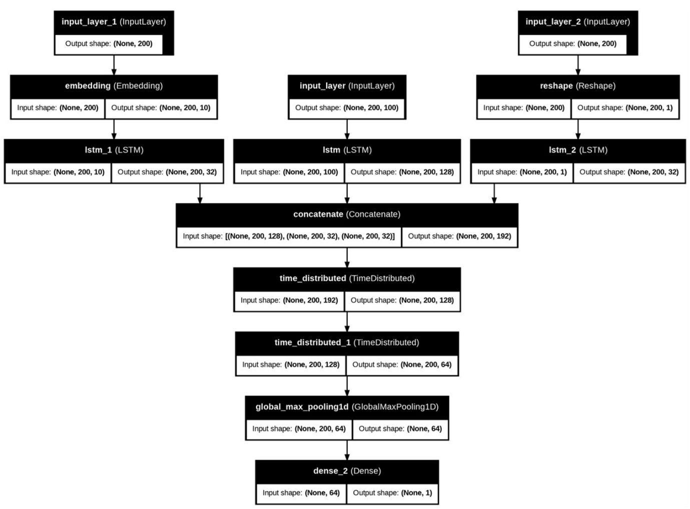

Fig. 2. Architecture of the LSTM Model

For the modeling phase, Long Short-Term Memory (LSTM) network were employed due to their abilities to handle sequential data, such as the textual messages in our dataset. The architecture of the model was designed to:

- Process the sequence of word embeddings from the messages.
- Capture long-term dependencies between words within the messages.

The model utilizes mutiple LSTM layers with the parameter *return_sequences*, which ensures that the LSTMs outputs the entire sequences for each time step, rather than just the final step. This is necessary because we want the complete sequence to be passed to further layers for processing.

Following the LSTM layers, we incorporate *TimeDistributed* layers, which allow the same dense layer to be applied to each time step of the sequence independently. This enables the model to process each time step of the input sequence and maintain the temporal structure.

The *GlobalMaxPooling1D* layer is used, which performs a max pooling operation across the different time steps of the sequence. This layer extracts the most significant feature from each sequence, reducing the dimensionality and focusing the model on the most relevant information for the prediction task.

The model was trained on the training set, and its performance was initially evaluated on the test set using metrics such as accuracy, precision, and recall.

In addition to the LSTM network, we explored another deep learning model to perform the classification task and compare their performance. Specifically, we implemented a Gated Recurrent Unit network. The architecture is shown in the Figure 3.

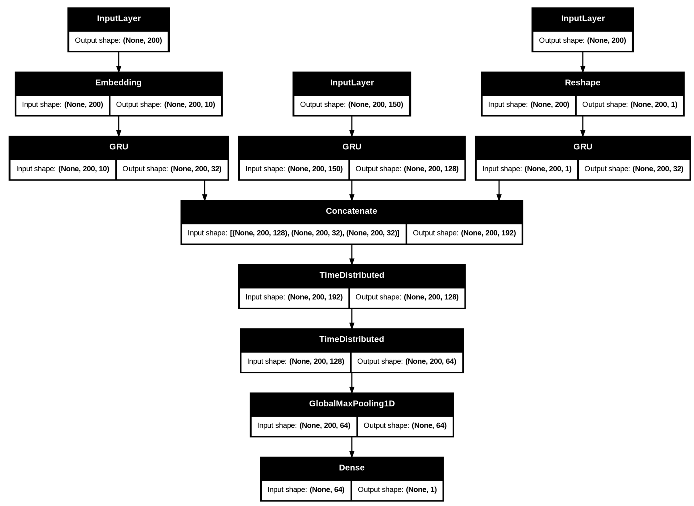

Fig. 3. Architecture of GRU Model

On the other side, we applied Decision Trees and Support Vector Machines (SVM) to the vectorized data. These models, while not inherently designed for sequential data, offer alternative approaches to classification and provide valuable baseline comparisons. A concatenation of the vectors is performed in order to pass it through the models. By evaluating the results across these diverse models, we aim to identify the best-performing approach for our dataset.

The further evaluation was conducted using data from another period (with the same format but not part of the training set) to assess the generalizability of the model. The best model from was test on the entire dataset from this period and on a subset comprising different percentages of the data to analyze the prediction results under different data volumes.

V. RESULTS

The performance of different models under various Word2Vec configurations, as shown in Table I, highlights key insights regarding the comparative effectiveness of deep learning and traditional machine learning approaches.

The results indicate that deep learning models, specifically LSTM and GRU, outperform traditional machine learning models across all configurations. Among these, the **LSTM model** with a Word2Vec configuration of vector size 100 and window size 7 achieved the highest overall performance. It recorded a training accuracy of **0.975** and a testing accuracy of **0.929**, with a testing F1 Score of **0.929**. This makes LSTM the best-performing model in this study. The GRU model also showed competitive performance, particularly with the same configuration, achieving a testing accuracy of **0.919** and an F1 Score of **0.918**, slightly behind LSTM.

For traditional machine learning models, Decision Trees (DT) demonstrated better performance compared to Support Vector Machines (SVM). The **Decision Tree model** with a Word2Vec configuration of vector size 150 and window size 7 was the best among traditional models. It achieved a testing accuracy of **0.789** and an F1 Score of **0.791**. In contrast, the SVM model consistently underperformed across all configurations, particularly for larger vector sizes.

The LSTM model, utilizing Word2Vec with a vector size of 100 and a window size of 7, demonstrated the best performance. The prediction results obtained with this model are presented in Figure 4. To further analyze its performance, we evaluated the model using varying percentages of data.

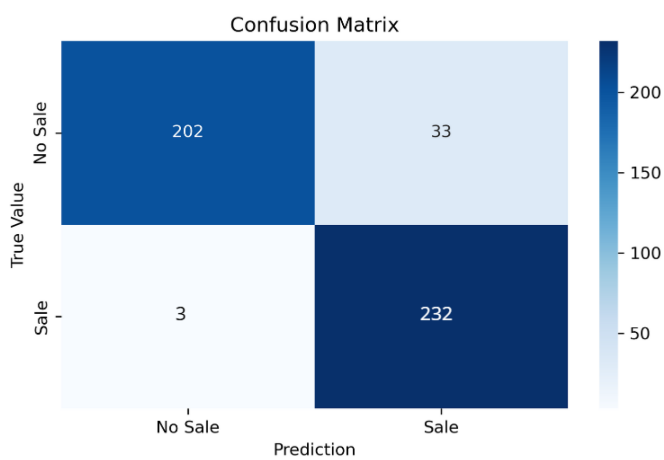

Fig. 4. Confusion Matrix of the best LSTM-Model

The accuracys obtained with different percentages of data are shown in the figure 5

VI. Discussion

The findings from this study demonstrate the effectiveness of deep learning models in handling the dataset for classification. The LSTM model with a Word2Vec configuration of vector size 100 and window size 7 emerged as the top-performing model, achieving a testing accuracy of 92.9% and an F1 Score of 92.9%. This performance underscores the LSTM's capacity to capture long-term dependencies and

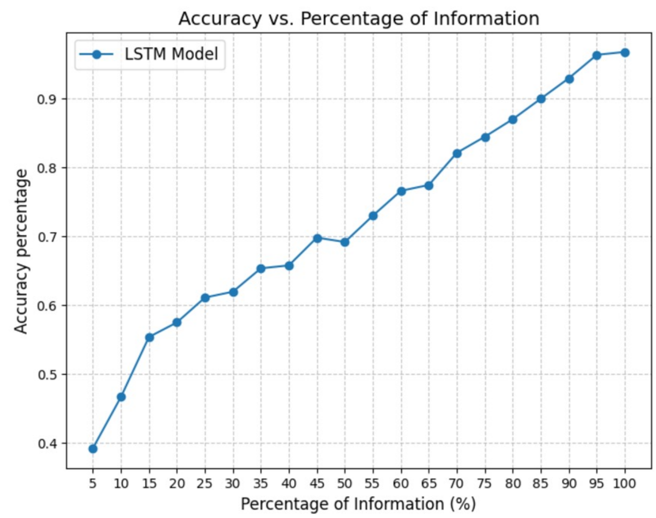

Fig. 5. Accuracys of the best LSTM-Model with different percentages of data

sequential context within the data, making it particularly suited for text classification tasks.

The GRU model also performed well, achieving competitive results with a testing accuracy of 91.9% and an F1 Score of 91.8%. This demonstrates that GRUs, as a simpler and computationally efficient alternative to LSTMs, can provide robust performance in similar tasks, although with less precision when it comes to capturing intricate dependencies.

Traditional machine learning models, such as Decision Trees and Support Vector Machines, struggled to match the performance of deep learning models. The Decision Tree model, with its best configuration, with Word2Vec vector size 150 and window size 7, achieved a testing accuracy of 78.9% and an F1 Score of 79.1%, outperforming SVM but still falling significantly short of the deep learning models. SVM consistently underperformed, indicating its limitations in handling high-dimensional and sequential data.

The progressive evaluation of the LSTM model using different percentages of data provided further insights into its robustness and scalability. As shown in Figure 5, the accuracy of the model increased steadily with the amount of data used, reaching a peak accuracy of 97% with the full dataset. This trend highlights the importance of data volume in improving model performance, particularly for deep learning approaches. The confusion matrix (Figure 4) further validates the LSTM model's ability to accurately classify messages, with minimal misclassifications.

These results also emphasize the critical role of preprocessing and dataset preparation. The steps taken to clean, tokenize, and standardize the text, as well as to filter out short and incomplete conversations, likely contributed to the model's ability to generalize effectively. Additionally, the decision to use the last 200 messages per conversation ensured consistency in the data and avoided overfitting to unusually long conversations.

Finally, the Word2Vec configurations played a pivotal role

TABLE I
TRAINING AND TESTING RESULTS FOR DIFFERENT MODELS AND WORD2VEC CONFIGURATIONS.

W2V Configuration	Model	Training			Testing		
		Accuracy	Precision	F1 Score	Accuracy	Precision	F1 Score
100, window 5	LSTM	0.969	0.969	0.969	0.912	0.922	0.912
	GRU	0.961	0.962	0.961	0.914	0.921	0.914
	DT	0.856	0.867	0.854	0.800	0.819	0.797
	SVM	0.836	0.836	0.836	0.766	0.766	0.766
100, window 7	LSTM	**0.975**	**0.975**	**0.975**	**0.929**	**0.932**	**0.929**
	GRU	0.973	0.973	0.973	0.919	0.923	0.918
	DT	0.856	0.857	0.856	0.796	0.799	0.795
	SVM	0.842	0.843	0.842	0.770	0.770	0.770
150, window 5	LSTM	0.964	0.965	0.964	0.938	0.940	0.938
	GRU	0.961	0.961	0.961	0.916	0.924	0.916
	DT	0.857	0.859	0.856	0.778	0.785	0.777
	SVM	0.843	0.843	0.843	0.776	0.777	0.776
150, window 7	LSTM	0.961	0.961	0.961	0.919	0.918	0.919
	GRU	0.961	0.961	0.961	0.903	0.920	0.902
	DT	0.863	0.863	0.863	0.789	0.792	0.791
	SVM	0.843	0.843	0.843	0.783	0.783	0.783

in determining model performance. Configurations with a window size of 7 consistently outperformed those with a window size of 5, suggesting that a broader contextual window provides richer representations of word relationships. This aligns with the nature of conversational data, where understanding context often requires a wider perspective.

VII. CONCLUSIONS

This study set out to address the challenge of optimizing sales conversations through predictive analytics, using NLP techniques to extract actionable insights from sequential text data. The results demonstrate that deep learning models, particularly Long Short-Term Memory (LSTM) networks, are highly effective for this purpose. With an F1 score of 0.929, the LSTM model trained on Word2Vec embeddings consistently outperformed both traditional machine learning approaches and alternative neural architectures like GRUs. The superior performance of deep learning models underscores their capacity to capture complex dependencies and patterns in conversational data.

From a practical perspective, these findings are highly relevant for businesses aiming to enhance their customer interactions and sales processes. The ability to predict sales outcomes during conversations could empower sales consultants with real-time guidance, enabling them to focus on high-propensity opportunities and refine their engagement strategies. Moreover, companies can use the insights to train their staff and fine-tune their conversational approaches based on data-driven patterns.

Despite these promising outcomes, the study also highlights several areas for future work. One limitation lies in the use of a single-language dataset. Extending the model to multi-lingual and cross-cultural datasets would make it more universally applicable.

Another promising direction involves implementing the model for real-time analysis, allowing sales teams to adjust their approaches dynamically during ongoing conversations. Exploring the integration of additional features, such as customer demographics or prior purchase history, could also yield richer predictive insights.

In conclusion, this research confirms the potential of NLP and machine learning to revolutionize sales optimization. By embracing these technologies, businesses can not only improve operational efficiency but also build stronger, more personalized connections with their customers, ensuring sustained success in an increasingly digital marketplace.

REFERENCES

[1] S. Y. Alwan, Y. Hu, A. A. M. H. Al Asbahi, Y. K. Al Harazi, and A. K. Al Harazi, "Sustainable and resilient e-commerce under COVID-19 pandemic: a hybrid grey decision-making approach," Environmental Science and Pollution Research, vol. 30, no. 16, pp. 47328-47348, 2023.

[2] United Nations Conference on Trade and Development, "Global e-commerce jumps to $26.7 trillion, COVID-19 boosts online sales," UNCTAD, May 2021. Available: https://unctad.org/news/global-e-commerce-jumps-267-trillion-covid-19-boosts-online-sales

[3] C. Osorio Chávez, "Análisis de la percepción del servicio de los clientes de 18 a 35 años sobre el comercio electrónico de Falabella en el sector moda, Arequipa (2020-2021)," 2023.

[4] Y. Jiang and N. Stylos, "Triggers of consumers' enhanced digital engagement and the role of digital technologies in transforming the retail ecosystem during COVID-19 pandemic," *Technological Forecasting and Social Change*, vol. 172, p. 121029, 2021.

[5] J. Hill, "The Handbook of Customer Satisfaction and Loyalty Mesurement". 2006

[6] G. Cardeño Agudelo, "Análisis de la capacidad de atención en el call center de Alkomprar," 2020.

[7] K. Chowdhary and K. R. Chowdhary, "Natural language processing", Fundamentals of Artificial Intelligence, pp. 603-649, 2020.

[8] Y. HaCohen-Kerner, D. Miller, and Y. Yigal, "The influence of preprocessing on text classification using a bag-of-words representation," PloS One, vol. 15, no. 5, pp. e0232525, May 2020.

[9] P. A. Olujimi and A. Ade-Ibijola, "NLP techniques for automating responses to customer queries: A systematic review," Discover Artificial Intelligence, vol. 3, pp. 20, Apr. 2023.

[10] J. Bakker, "Predicting Shopping Intention on Sequential Data Using Different Recurrent Neural Network Algorithms," Master's thesis, Tilburg University, Dept. of Cognitive Science and Artificial Intelligence, Tilburg, Netherlands, Dec. 2020.

[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[12] M. A. Alshamari, "Evaluating user satisfaction using deep-learning-based sentiment analysis for social media data in Saudi Arabia's telecommunication sector," Computers, vol. 12, no. 9, pp. 170, Aug. 2023.

[13] U. Singh, A. Saraswat, H. K. Azad, K. Abhishek, and S. Shitharth, "Towards improving e-commerce customer review analysis for sentiment detection," Scientific Reports, vol. 12, no. 21983, pp. 1-15, Dec. 2022.
[14] Y. Piris and A.-C. Gay, "Customer satisfaction and natural language processing," Journal of Business Research, vol. 124, pp. 264–271, Dec. 2020.
[15] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, *Distributed Representations of Words and Phrases and their Compositionality*, *Advances in Neural Information Processing Systems*, 2013.
[16] S. Hochreiter and J. Schmidhuber, *Long Short-Term Memory*, *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
[17] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*, *arXiv preprint arXiv:1406.1078*, 2014.
[18] L. Breiman, *Classification and Regression Trees*, *Wadsworth International Group*, 1984.
[19] C. Cortes and V. Vapnik, *Support-Vector Networks*, *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.