




Sentiment Analysis of Customers in Iran's Automotive Industry Using the LSTM Method

Lida. Shojaei Barjouei¹, Ghasem. Bakhshandeh^{1*}, Mahmood. Daniali Deh houz²

¹ Department of Management, Ahv.C., Islamic Azad University, Ahvaz, Iran

² Department of Accounting, Iz.C., Islamic Azad University, Izeh, Iran

* Corresponding author email address: gh.bakhshandeh@iau.ac.ir

Article Info

Article type:

Original Research

How to cite this article:

Shojaei Barjouei, L., Bakhshandeh, G., & Daniali Deh houz, M. (2026). Sentiment Analysis of Customers in Iran's Automotive Industry Using the LSTM Method. *Journal of Resource Management and Decision Engineering*, 5(6), 1-13.

<https://doi.org/10.61838/kman.jrmde.319>



© 2026 the authors. Published by KMAN Publication Inc. (KMANPUB). This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

ABSTRACT

Given the large volume of customer opinions and complaints regarding automobiles in Iranian virtual platforms, this study aimed to develop an accurate sentiment analysis model based on Long Short-Term Memory (LSTM) neural networks for processing Persian texts. The primary objective was to improve the accuracy of sentiment classification (positive, negative, and neutral) by overcoming the challenges associated with imbalanced data and linguistic limitations. Data were collected from 12 valid online sources, including 4,348 customer comments regarding various automobile models. Three LSTM architectures (single-layer, two-layer, and three-layer) were implemented and compared using fixed parameters (Embedding = 128, Hidden Units = 64, Dropout = 0.3). To enhance performance, Data Augmentation, the SMOTE algorithm, and FastText word embeddings were employed. The three-layer LSTM model outperformed the shallower models, achieving a validation accuracy of 62.01%, which increased to 95.45% after applying optimization techniques. The two-layer model demonstrated superior performance in identifying positive comments, whereas the three-layer model showed a more balanced performance across all sentiment classes. The issue of overfitting in simpler models was mitigated through the addition of layers and the use of Dropout. The findings indicated that the two-layer model improved overall accuracy, particularly in recognizing positive sentiments, compared to the single-layer model. Furthermore, the three-layer LSTM model demonstrated greater balance and robustness across all classes. To further improve performance, the three-layer model was retrained using Data Augmentation techniques and the SMOTE algorithm for data balancing. The modified version substantially improved overall accuracy and validation performance.

Keywords: Sentiment analysis, automotive industry customers, Iran, LSTM, SMOTE, Data Augmentation

1. Introduction

The rapid expansion of digital communication technologies and social media platforms has fundamentally transformed the way organizations interact with customers and evaluate consumer behavior. In contemporary digital environments, customers continuously share their experiences, opinions, and evaluations regarding products and services through online reviews, social media posts, forums, and specialized platforms. These textual interactions have created a vast volume of unstructured data that contains valuable insights into customer satisfaction, emotional responses, purchasing intentions, and behavioral patterns. As a result, sentiment analysis has emerged as one of the most important applications of artificial intelligence, machine learning, and natural language processing for extracting meaningful information from textual data (Birjali et al., 2021; Mao et al., 2024). Sentiment analysis enables organizations to automatically identify customers' emotional orientations toward products and services and supports data-driven decision-making processes in marketing, product development, and customer relationship management (Hossain, 2024; Selimovic et al., 2024).

In recent years, the increasing availability of online customer-generated content has significantly expanded the importance of text mining and sentiment analysis methods across different industries. Organizations increasingly rely on computational models to analyze consumer opinions because manual processing of large-scale textual data is costly, time-consuming, and often impractical. Advances in machine learning and deep learning techniques have improved the ability of computational systems to understand contextual semantics, hidden emotional patterns, and linguistic dependencies within textual data (Mao et al., 2024; Rodriguez-Ibanez et al., 2023). Consequently, sentiment analysis has become an essential analytical tool for evaluating customer attitudes, predicting market trends, and identifying competitive advantages in digital markets (Mehrkian & Davari-Ardakani, 2025; Ricchiuti & Sperli, 2025).

The automotive industry represents one of the sectors in which customer opinions play a highly influential role in shaping market reputation, consumer trust, and product competitiveness. Automobile customers frequently express their perceptions regarding vehicle quality, fuel consumption, safety, design, after-sales services, technical performance, and maintenance costs through online platforms and automotive review websites. These opinions

significantly influence purchasing decisions and consumer preferences in competitive markets. Therefore, analyzing customer sentiments in the automotive sector provides valuable information for manufacturers, distributors, and policymakers seeking to improve product quality and customer satisfaction (Lin & Jen, 2024; Wen et al., 2024). The growing importance of electric vehicles, digital automotive ecosystems, and online purchasing behaviors has further intensified the need for accurate sentiment analysis systems capable of processing automotive-related textual data (Prasastio et al., 2024).

Traditional sentiment analysis methods primarily relied on lexicon-based approaches and classical machine learning algorithms such as Naïve Bayes, Support Vector Machines, and Decision Trees. Although these approaches demonstrated acceptable performance in structured datasets, they often struggled to capture complex semantic relationships, contextual meanings, and long-term dependencies in natural language texts (Birjali et al., 2021; Wang & Zheng, 2016). Moreover, traditional approaches are highly dependent on manual feature engineering and predefined dictionaries, which limits their ability to generalize across different linguistic domains and contexts. These limitations become even more pronounced in social media environments where texts are highly informal, noisy, and linguistically inconsistent (Farajzadeh et al., 2023; Rodriguez-Ibanez et al., 2023).

The emergence of deep learning techniques has substantially improved the effectiveness of sentiment analysis systems by enabling automatic feature extraction and contextual representation learning. Deep neural networks can model nonlinear relationships and hierarchical semantic structures within textual data more effectively than conventional machine learning approaches (Karimi Dastgerdi & Zamani Boroujeni, 2020; Mirzaee et al., 2024). Among deep learning architectures, recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks have gained considerable attention due to their ability to process sequential data and preserve contextual dependencies across long textual sequences (Amjadian, 2024; Dam et al., 2025). LSTM networks are particularly effective in sentiment analysis tasks because they can mitigate the vanishing gradient problem associated with conventional RNNs and maintain long-term contextual memory during text processing (Dam et al., 2025; Ricchiuti & Sperli, 2025).

LSTM-based architectures have been successfully applied in various domains including finance, tourism,

healthcare, digital marketing, and product recommendation systems. Research has demonstrated that LSTM models can achieve superior predictive performance in identifying emotional patterns and customer sentiments compared with traditional machine learning methods (Mehrkian & Davari-Ardakani, 2025; Puh & Bagic Babac, 2023). In the financial sector, LSTM-based sentiment analysis models have been used to predict stock market trends by integrating quantitative indicators with textual sentiment information extracted from online sources (Mehrkian & Davari-Ardakani, 2025; Ricchiuti & Sperli, 2025). Similarly, in digital marketing environments, deep learning approaches have improved customer behavior prediction and personalized recommendation systems through more accurate interpretation of consumer sentiments (Hossain, 2024; Mirzaee et al., 2024).

Despite these advancements, sentiment analysis remains a challenging task, particularly in multilingual and low-resource language environments such as Persian. Persian textual data present several linguistic complexities including orthographic inconsistencies, informal writing styles, colloquial expressions, and the absence of large-scale annotated corpora. These characteristics complicate the preprocessing, tokenization, and semantic interpretation of Persian texts (Kazemi et al., 2023; Mohammadi, 2023). Furthermore, social media texts in Persian frequently contain abbreviations, emojis, spelling variations, and mixed linguistic structures that reduce the effectiveness of traditional natural language processing methods (Farajzadeh et al., 2023; Mohammadi, 2023). Consequently, developing robust deep learning frameworks specifically designed for Persian sentiment analysis remains an important research priority.

Another major challenge in sentiment analysis research involves the issue of imbalanced datasets. In many real-world applications, positive sentiments are significantly more frequent than neutral or negative opinions, resulting in biased model performance and poor classification accuracy for minority classes. Imbalanced datasets can reduce the generalization capability of machine learning models and increase the risk of overfitting toward dominant sentiment categories (Birjali et al., 2021; Mao et al., 2024). To address this issue, researchers have increasingly incorporated data balancing techniques such as the Synthetic Minority Over-sampling Technique (SMOTE), data augmentation strategies, and weighted loss functions into deep learning frameworks (Dam et al., 2025; Kazemi et al., 2023). These methods enhance model robustness by increasing the

representation of underrepresented classes and improving semantic diversity within training datasets.

Recent developments in artificial intelligence have also introduced transformer-based architectures and large language models such as BERT and ChatGPT into sentiment analysis research. Studies have shown that these advanced models can improve contextual understanding and semantic representation in textual classification tasks (Sudirjo et al., 2023; Susnjak, 2024). However, transformer-based systems often require substantial computational resources, large annotated datasets, and extensive training procedures, which may limit their applicability in specialized or low-resource domains. In contrast, LSTM architectures remain computationally efficient and highly effective for sequential text classification tasks, especially when combined with optimized preprocessing techniques and word embedding models (Amjadian, 2024; Dam et al., 2025).

Word embedding techniques also play a critical role in enhancing sentiment analysis performance. Embedding models such as Word2Vec and FastText transform textual units into dense numerical vectors that capture semantic similarities and contextual relationships among words. FastText embeddings are particularly advantageous in Persian language processing because they consider subword information and can better represent rare or morphologically complex words (Huang et al., 2022; Mirzaee et al., 2024). The integration of embedding models with LSTM architectures has significantly improved the accuracy and contextual sensitivity of sentiment analysis systems in multilingual environments (Lin & Jen, 2024; Wen et al., 2024).

The automotive market in Iran presents a particularly valuable context for sentiment analysis research due to the high level of public engagement in online automotive discussions. Iranian customers actively share detailed opinions regarding vehicle reliability, maintenance costs, fuel efficiency, design quality, and customer services on automotive websites and social media platforms. These online discussions provide a rich source of unstructured data for analyzing consumer attitudes and identifying market trends. Nevertheless, relatively limited research has focused specifically on Persian-language sentiment analysis within the Iranian automotive sector using advanced deep learning approaches (Farajzadeh et al., 2023; Mohammadi, 2023). Existing studies have primarily concentrated on general social media sentiment analysis, financial forecasting, or urban perception modeling rather than domain-specific

automotive applications (Kazemi et al., 2023; Mehrkian & Davari-Ardakani, 2025).

Furthermore, many previous studies have focused on binary sentiment classification and have not adequately addressed the complexity of multi-class sentiment analysis involving positive, negative, and neutral categories simultaneously. Multi-class sentiment analysis is considerably more challenging because neutral sentiments often contain ambiguous emotional structures and contextual nuances that are difficult to classify accurately (Huang et al., 2022; Wang & Zheng, 2016). Therefore, developing optimized LSTM architectures capable of achieving balanced performance across multiple sentiment classes remains an important methodological challenge.

The present study contributes to the existing literature by developing and evaluating single-layer, two-layer, and three-layer LSTM architectures for Persian sentiment analysis in the Iranian automotive industry. In addition, the study incorporates preprocessing optimization, FastText word embeddings, Data Augmentation techniques, and the SMOTE algorithm to improve classification accuracy and reduce the effects of class imbalance. By comparing different LSTM architectures and evaluating their predictive performance across multiple sentiment categories, this research provides a comprehensive framework for deep learning-based sentiment analysis in Persian-language automotive datasets. The findings of this study may support automotive companies, marketers, and policymakers in understanding customer perceptions and improving strategic decision-making processes in digital environments (Selimovic et al., 2024; Wen et al., 2024).

Therefore, the aim of the present study was to analyze customer sentiments in the Iranian automotive industry using LSTM-based deep learning models and to identify the most effective architecture for classifying Persian-language customer reviews into positive, negative, and neutral sentiment categories.

2. Methods and Materials

This study was designed as an applied quantitative research project in the field of intelligent text analytics and sentiment analysis using deep learning methods. The research focused on developing and evaluating Long Short-Term Memory (LSTM) neural network architectures for sentiment classification of Persian-language customer opinions in the Iranian automotive industry. The study adopted a comparative experimental design in which

different LSTM configurations were implemented and evaluated to determine the most efficient architecture for classifying customer sentiments into positive, negative, and neutral categories. The research process involved large-scale extraction of textual data from online automotive platforms, preprocessing of Persian-language texts, training deep learning models, and performance evaluation using standard classification metrics.

The dataset used in this research consisted of customer reviews and comments collected from reputable Iranian automotive websites and online platforms. The textual data were extracted from multiple online sources, including Khodrobank.com, 1car.ir, Hamrah-mechanic.com, Asbe-bokhar.com, Pedal.ir, Mashin3.com, Bama.ir, Bitrun.info, Becharkh.com, Nazarkade.com, Carbalad.com, Pardiskhodro.com, and Khodro45.com. These platforms were selected because of their extensive user engagement and the high volume of customer-generated content related to automobiles available in the Iranian market. The final dataset included 8,696 customer comments categorized into positive, negative, and neutral sentiment classes. Among these comments, 4,318 were labeled as positive, 3,284 as negative, and 1,094 as neutral. The sentiment labels were assigned according to the semantic orientation and emotional polarity expressed within the textual reviews.

In order to improve the representativeness and diversity of the dataset, comments were collected from different automobile categories and vehicle models commonly available in the Iranian automotive market. The dataset included reviews related to various domestic and imported vehicles such as X22 Pro, H30 Cross, Atlas, MVM X33, Peugeot 207, Peugeot Pars, Tara, JAC S3, JAC S5, Dena, Rana, Rana Plus, Saina, Samand Soren, Sahand, Shahin, and Quick. The inclusion of multiple vehicle categories enabled the study to capture a broader range of customer experiences, linguistic expressions, and sentiment patterns associated with automotive products and services in Iran. This diversity contributed to enhancing the generalizability and robustness of the developed sentiment analysis models.

The initial stage of the research involved the systematic collection of customer-generated textual data from online automotive platforms. Data were extracted from user comments, product reviews, customer evaluations, and discussion posts related to different automobile models. Prior to the data extraction process, specific search parameters and filtering criteria were defined to identify relevant customer opinions and automotive-related discussions. The collected texts reflected customers'

perceptions regarding vehicle quality, technical performance, fuel consumption, design, comfort, after-sales services, maintenance costs, and overall user satisfaction.

After data collection, extensive preprocessing procedures were conducted to prepare the textual data for deep learning analysis. Since Persian-language textual data often contain linguistic inconsistencies and informal writing patterns, several preprocessing operations were applied to improve data quality and model performance. The preprocessing stage included textual noise removal through the elimination of unnecessary punctuation marks, redundant numbers, URLs, hyperlinks, and emojis. In addition, orthographic normalization was performed to standardize common spelling variations and convert semantically identical Persian characters into unified forms. This step was particularly important because Persian online texts frequently contain inconsistencies in character representation and word spelling.

Subsequently, tokenization procedures were applied to segment sentences into individual lexical units suitable for computational processing. Persian stop words with limited semantic value were removed using a predefined Persian stop-word list in order to reduce dimensionality and improve semantic representation. Following the cleaning and normalization stages, textual data were transformed into numerical vector representations using pre-trained Word2Vec embeddings trained on Persian corpora with a dimensionality of 300 features. The use of distributed word representations enabled the models to capture semantic relationships and contextual similarities among words more effectively than traditional feature extraction techniques.

To address the issue of class imbalance and improve classification performance, data balancing and augmentation techniques were incorporated into the modeling process. The Synthetic Minority Over-sampling Technique (SMOTE) was employed to generate synthetic samples for minority classes and reduce distributional imbalance among sentiment categories. Furthermore, Data Augmentation methods were utilized to increase data diversity and enhance the generalization capability of the models. These augmentation procedures contributed to improving model robustness, reducing overfitting, and increasing classification accuracy, particularly for underrepresented sentiment classes.

The present study implemented three different LSTM architectures, including single-layer, two-layer, and multi-layer LSTM networks. The single-layer architecture was designed for relatively simple sequential learning tasks and

short textual dependencies. The two-layer architecture provided a balance between computational complexity and predictive performance while enabling better extraction of temporal and semantic patterns. The multi-layer LSTM architecture, consisting of three or more hidden layers, was designed for more complex sequential learning and long-term contextual dependency modeling. The LSTM models were selected because of their superior ability to capture long-range dependencies and contextual information in textual sequences compared with traditional recurrent neural networks.

Each LSTM cell consisted of three primary gating mechanisms, including the input gate, forget gate, and output gate. The input gate controlled the extent to which new information entered the memory cell, whereas the forget gate determined which irrelevant information should be discarded from memory. The output gate regulated the transfer of processed information to subsequent layers and output units. These gating mechanisms enabled the LSTM networks to effectively preserve long-term dependencies while mitigating the vanishing gradient problem commonly observed in traditional recurrent neural network architectures.

The implementation and evaluation of the deep learning models were performed using supervised machine learning procedures. The dataset was divided into training and testing subsets, with 80% of the data allocated for model training and 20% reserved for testing and evaluation purposes. In addition, 20% of the training data were used as a validation subset during the training process to monitor model performance and prevent overfitting. The models were trained using 30 epochs with a batch size of 64 observations per iteration. The Adam optimization algorithm was employed as the primary optimization technique with a learning rate of 0.001 due to its computational efficiency and adaptive learning capabilities.

The categorical cross-entropy function was used as the primary loss function for multi-class sentiment classification. To reduce overfitting and improve generalization performance, a Dropout rate of 0.3 was applied within the hidden layers of the LSTM networks. The inclusion of Dropout layers prevented excessive dependency on specific neurons and improved the stability of the learning process. Comparative analyses were conducted among the single-layer, two-layer, and multi-layer LSTM architectures to determine the optimal configuration for Persian sentiment classification in the automotive domain.

Model performance was evaluated using several standard classification metrics, including accuracy, recall, precision, and F1-score. Accuracy was used to measure the overall proportion of correctly classified observations, whereas precision assessed the proportion of correctly predicted positive observations relative to all predicted positive cases. Recall evaluated the ability of the model to correctly identify actual positive instances, and the F1-score provided a balanced measure combining both precision and recall. The comparative evaluation of these metrics enabled the identification of the most effective LSTM architecture in terms of classification performance, class balance, and generalization capability. The final model selection was based on the highest overall validation performance and balanced predictive accuracy across all sentiment categories.

Table 1

Distribution of Sentiments in the Single-Layer LSTM Model

Positive Comments	Negative Comments	Neutral Comments	Total Comments
487	313	200	1000

Descriptive analysis of the training data indicated that the average text length was 46.15 units with a standard deviation of 41.42, reflecting a relatively high degree of variability in textual sequence length. The shortest text contained only one unit, whereas the longest comment reached 251 units. In addition, 25% of the comments contained fewer than 16 units, while 75% contained fewer than 64 units. This distribution indicated the presence of highly diverse textual structures, including short, medium, and long comments, which provided suitable conditions for sentiment classification using deep learning models.

Table 2

Evaluation Metrics of the Single-Layer LSTM Sentiment Analysis Model

Metric	Numerical Value	Percentage (%)
Accuracy	0.4600	46.00
Misclassification Rate	0.5400	54.00
Weighted Precision	0.4423	44.23
Weighted Recall	0.4600	46.00
Weighted F1-Score	0.4427	44.27
Mean Specificity	0.8507	85.07

Further analysis of the classification report revealed that the model performed differently across sentiment categories. The positive sentiment class achieved the highest classification performance with a precision value of 0.55, recall value of 0.67, and F1-score of 0.61. In contrast, the

3. Findings and Results

The single-layer LSTM model was initially implemented using a dataset consisting of 1,000 selected customer comments extracted from Iranian automotive platforms. The dataset contained four primary columns, including record ID, customer comments, vehicle type, and sentiment type. The sentiment distribution in this dataset demonstrated a relatively imbalanced structure, with positive comments representing the majority of observations. Specifically, 487 comments were categorized as positive, 313 as negative, and 200 as neutral. The dataset was divided into training and testing subsets, including 750 training samples and 250 testing samples across three sentiment classes.

The evaluation results of the single-layer LSTM model demonstrated moderate predictive performance. The overall accuracy of the model reached 46.00%, while the misclassification rate was reported at 54.00%. The weighted precision, weighted recall, and weighted F1-score were 44.23%, 46.00%, and 44.27%, respectively. However, the average specificity of the model reached 85.07%, indicating a relatively strong ability to correctly identify non-target classes.

neutral class demonstrated weaker predictive performance, with an F1-score of only 0.29. The weighted average F1-score remained at 0.44, confirming the moderate overall capability of the model in sentiment prediction.

Table 3

Classification Report of the Single-Layer LSTM Model

Sentiment Class	Precision	Recall	F1-Score	Support
Neutral	0.30	0.28	0.29	50
Positive	0.55	0.67	0.61	92
Negative	0.44	0.46	0.45	63
Overall Accuracy	—	—	0.46	250
Macro Average	0.40	0.36	0.36	—
Weighted Average	0.44	0.46	0.44	—

The two-layer LSTM model was subsequently developed using 2,000 customer comments. The dataset included 956 positive comments, 709 negative comments, and 335 neutral comments. The dataset was divided into 1,500 training samples and 500 testing samples. Statistical analysis of the textual sequences indicated that the average text length was 42.6 units with a standard deviation of 41.7. The minimum sequence length was zero, indicating the existence of empty or extremely short comments, whereas the maximum sequence length reached 345 units. Quartile analysis demonstrated that 25% of comments contained fewer than 14 units, the median text length was 29 units, and 75% of comments contained fewer than 56 units. These findings highlighted the necessity of preprocessing operations such

as padding and truncation to standardize sequence lengths prior to model training.

The evaluation metrics of the two-layer LSTM model demonstrated noticeable improvement compared with the single-layer architecture. The overall model accuracy increased to 50.80%, while the weighted F1-score improved to 49.56%. The model showed particularly strong performance in detecting positive sentiments, achieving a recall value of 0.70 and precision value of 0.66 for the positive class. Nevertheless, weaker performance was still observed for minority classes, especially neutral sentiments. The specificity value remained high at 85.90%, confirming the model's ability to distinguish irrelevant classes effectively.

Table 4

Evaluation Metrics of the Two-Layer LSTM Sentiment Analysis Model

Metric	Numerical Value	Percentage (%)
Accuracy	0.5080	50.80
Misclassification Rate	0.4920	49.20
Weighted Precision	0.4890	48.90
Weighted Recall	0.5080	50.80
Weighted F1-Score	0.4956	49.56
Mean Specificity	0.8590	85.90

The classification report of the two-layer architecture further confirmed the superiority of this model over the simpler architecture. The positive sentiment class obtained the highest F1-score of 0.63, whereas the neutral class

remained relatively difficult to classify accurately. The weighted average precision, recall, and F1-score reached approximately 0.50, indicating moderate but improved classification performance.

Table 5

Classification Report of the Two-Layer LSTM Model

Sentiment Class	Precision	Recall	F1-Score	Support
Neutral	0.33	0.37	0.30	84
Positive	0.66	0.70	0.63	205
Negative	0.48	0.47	0.50	155
Overall Accuracy	0.51	—	—	500
Macro Average	0.35	0.35	0.36	—
Weighted Average	0.50	0.51	0.49	—

The three-layer LSTM model was then implemented using the complete dataset of 4,348 customer comments. This dataset consisted of 2,159 positive comments, 1,642 negative comments, and 547 neutral comments. The training dataset included 3,261 samples, while 1,087 samples were allocated to testing. Statistical analysis demonstrated that the average text length was 38.3 units with a standard deviation of 36.9. The shortest sequence length was zero and the maximum length was 316 units. Quartile distribution

analysis showed that 25% of comments contained fewer than 13 units, 50% contained fewer than 27 units, and 75% contained fewer than 51 units. These results indicated that the majority of comments were relatively short, although several long textual sequences were also present within the dataset. Consequently, preprocessing techniques such as padding and truncation were carefully implemented to optimize semantic feature extraction.

Figure 1

Confusion Matrix for the Testing Data of the Three-Layer LSTM Model

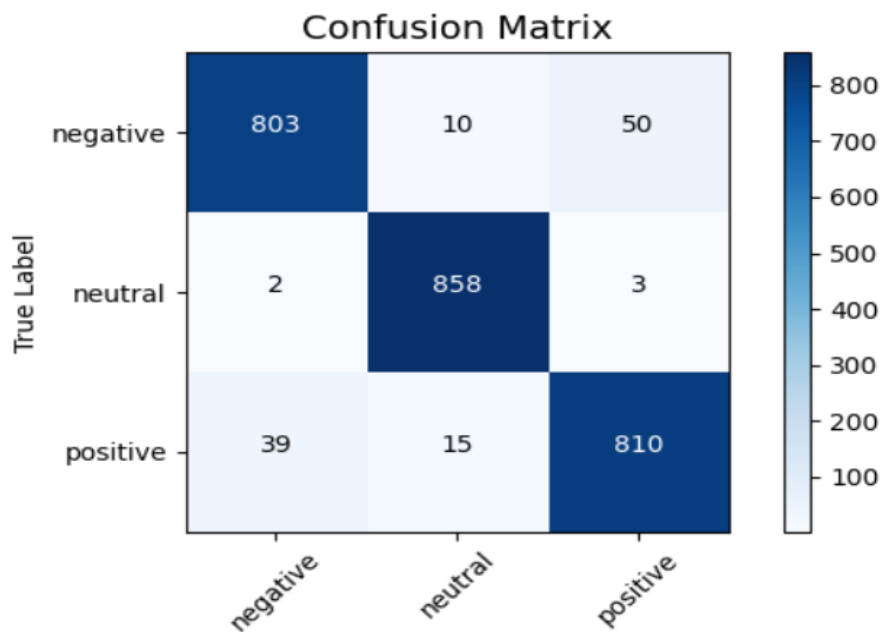
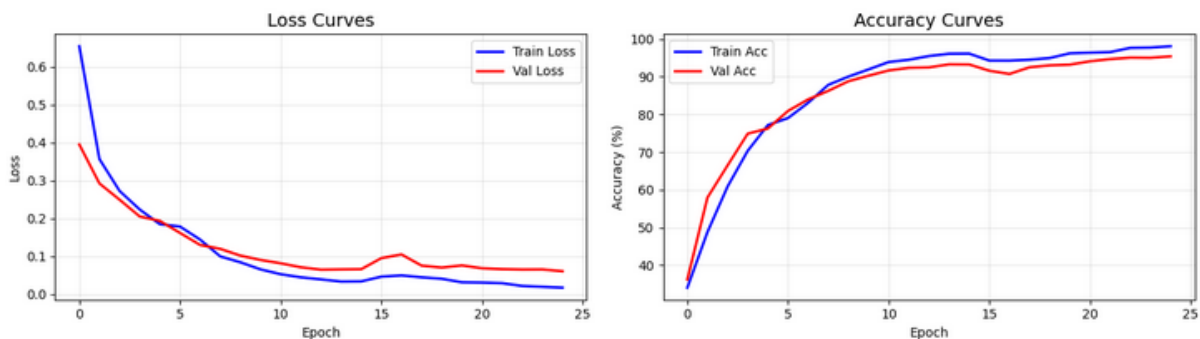


Figure 2

Training and Validation Accuracy and Loss Curves Across Epochs for the Three-Layer LSTM Model



The convergence patterns illustrated in the training and validation curves clearly demonstrated superior learning

stability and convergence behavior in comparison with the single-layer and two-layer architectures. The deeper

architecture showed improved semantic extraction capability and more stable optimization throughout the training process.

The evaluation metrics of the three-layer LSTM model indicated substantial improvement compared with the previous architectures. The overall classification accuracy

reached 65.28%, while the weighted precision, weighted recall, and weighted F1-score achieved 65.38%, 67.22%, and 65.13%, respectively. Furthermore, the model achieved a specificity value of 92.25%, indicating a strong ability to distinguish between sentiment categories accurately.

Table 6

Comparative Evaluation of the Proposed LSTM Architectures

Model	Best Training Accuracy	Validation Loss	Best Validation Accuracy	Epoch of Best Accuracy
1-Layer LSTM	99.07%	1.431	41.60%	Epoch 10
2-Layer LSTM	99.07%	1.272	53.80%	Epoch 13
3-Layer LSTM	97.70%	1.066	62.01%	Epoch 3
Modified Model	98.90%	0.842	95.45%	Epoch 8

The three-layer LSTM model demonstrated the highest performance among all baseline architectures, particularly in recognizing dominant sentiment categories such as positive sentiments with an F1-score of 0.70 and negative sentiments with an F1-score of 0.61. To further improve predictive capability, augmentation techniques and data balancing methods were applied to the three-layer architecture. The modified model achieved substantial improvements, with overall accuracy increasing to 95.52%, weighted precision reaching 95.50%, weighted recall reaching 95.50%, weighted F1-score reaching 95.50%, and specificity increasing to 96.00%. These findings demonstrate the effectiveness of combining deep LSTM architectures with augmentation and balancing techniques for Persian sentiment analysis tasks in the automotive domain.

In addition to the performance improvements, architectural comparisons demonstrated that all models utilized identical embedding dimensions of 128, hidden units of 64, and a dropout rate of 0.3 to prevent overfitting. However, the number of trainable parameters increased progressively from 178,117 parameters in the single-layer model to 244,677 parameters in the three-layer architecture. The study also employed pre-trained Persian FastText embeddings to preserve semantic relationships among rare and context-dependent Persian words. The implementation of Dropout regularization effectively reduced overfitting and improved model generalization performance across all experimental configurations.

4. Discussion and Conclusion

The present study aimed to develop and evaluate LSTM-based deep learning architectures for sentiment analysis of Persian-language customer comments in the Iranian

automotive industry. The findings demonstrated that increasing the depth of LSTM architectures substantially improved sentiment classification performance. The single-layer LSTM model achieved an overall accuracy of 46.00%, whereas the two-layer architecture improved the accuracy to 50.80%. The highest performance among the baseline architectures was observed in the three-layer LSTM model, which achieved an accuracy of 65.28% and a weighted F1-score of 65.13%. Furthermore, after applying Data Augmentation techniques and the SMOTE algorithm, the modified three-layer model achieved a significant increase in classification accuracy, reaching 95.52%. These findings indicate that deeper LSTM architectures combined with balancing and augmentation strategies can substantially improve the effectiveness of Persian sentiment analysis systems in specialized industrial domains.

One of the major findings of this study was the superiority of the multi-layer LSTM architecture over simpler configurations. The results demonstrated that the single-layer model exhibited limited ability to capture complex semantic relationships and contextual dependencies within customer reviews. This outcome is consistent with previous studies emphasizing that shallow neural architectures often struggle to represent long-term dependencies in textual sequences, particularly in noisy social media data (Birjali et al., 2021; Wang & Zheng, 2016). The improved performance of the two-layer and three-layer architectures confirms that deeper recurrent networks are more capable of extracting hierarchical semantic features and contextual emotional patterns from unstructured textual data (Dam et al., 2025; Ricchiuti & Sperli, 2025). The findings therefore support the argument that increasing architectural depth enhances

sequential learning capacity and contextual understanding in sentiment classification tasks.

The performance improvement observed in the three-layer LSTM model can also be explained by the enhanced ability of deep recurrent structures to preserve long-term contextual memory. Customer reviews in the automotive industry often contain compound opinions, contextual references, and mixed emotional expressions that require sequential semantic interpretation. The three-layer architecture demonstrated stronger convergence behavior and lower validation loss compared with the shallower models, suggesting that deeper networks can more effectively capture latent semantic structures in Persian texts. Similar findings have been reported in previous research demonstrating that LSTM architectures outperform conventional machine learning approaches in textual sentiment analysis because of their ability to model temporal dependencies and contextual relationships (Mao et al., 2024; Rodriguez-Ibanez et al., 2023). The present findings are also aligned with studies showing that multi-layer recurrent networks improve text classification accuracy in large-scale customer review datasets (Dam et al., 2025; Prasastio et al., 2024).

Another important finding of this study concerns the role of Data Augmentation and SMOTE techniques in improving classification performance. The modified three-layer model achieved remarkable improvements in accuracy, weighted precision, weighted recall, and weighted F1-score after balancing the dataset and augmenting textual samples. These findings indicate that data imbalance represented a major challenge in the original dataset, particularly because positive comments significantly outnumbered neutral sentiments. Imbalanced datasets frequently bias deep learning models toward dominant classes and reduce predictive accuracy for minority categories. The present results support previous studies emphasizing the effectiveness of balancing techniques and augmentation strategies in improving sentiment classification performance (Birjali et al., 2021; Kazemi et al., 2023). In addition, these findings confirm that synthetic sample generation methods such as SMOTE can improve model generalization capability and reduce overfitting in sentiment analysis tasks.

The findings also demonstrated that the positive sentiment class achieved the highest classification performance across all LSTM architectures. Both the two-layer and three-layer models showed higher precision and recall values for positive comments compared with neutral and negative classes. This outcome may be associated with

the larger representation of positive comments in the dataset, which enabled the models to learn positive semantic patterns more effectively. Similar observations have been reported in previous sentiment analysis studies where majority classes generally achieve higher classification accuracy because of increased exposure during model training (Mao et al., 2024; Wang & Zheng, 2016). In contrast, neutral comments demonstrated the weakest classification performance in most models. Neutral sentiments are often linguistically ambiguous and contain fewer emotionally explicit expressions, making them inherently more difficult to classify accurately (Huang et al., 2022; Rodriguez-Ibanez et al., 2023). The present findings therefore reinforce the methodological challenge of multi-class sentiment analysis in natural language processing research.

The high specificity values observed across all models represent another important result of this study. Even the weaker single-layer architecture achieved specificity values above 85%, while the modified model reached a specificity value of 96%. This finding indicates that the proposed models demonstrated strong capability in distinguishing irrelevant sentiment categories and minimizing false-positive classifications. High specificity is particularly valuable in customer sentiment analysis because inaccurate classification of customer dissatisfaction may lead to misleading managerial decisions and ineffective marketing strategies. The findings are consistent with studies emphasizing that deep learning models are highly effective in identifying class boundaries and reducing classification errors in customer opinion mining applications (Selimovic et al., 2024; Wen et al., 2024).

The findings of this study also highlight the importance of preprocessing techniques in Persian-language sentiment analysis. The dataset contained substantial variability in text length, spelling patterns, and linguistic structures. Consequently, preprocessing procedures including normalization, tokenization, stop-word removal, padding, truncation, and vector representation played a critical role in improving model performance. The successful application of FastText embeddings in this study confirms the importance of embedding-based semantic representation methods in multilingual sentiment analysis. FastText embeddings are particularly effective in Persian language processing because they preserve subword information and capture semantic relationships among morphologically complex words. Similar findings have been reported in previous research emphasizing that embedding-based representations significantly improve contextual

understanding and classification accuracy in deep learning frameworks (Lin & Jen, 2024; Mirzaee et al., 2024).

Another notable result of this study was the reduction of overfitting through the use of Dropout regularization and deeper architectures. The single-layer model demonstrated extremely high training accuracy but substantially lower validation accuracy, indicating poor generalization performance and overfitting to the training data. In contrast, the deeper architectures exhibited improved convergence behavior and more balanced training-validation performance. This finding supports previous research indicating that Dropout layers and architectural optimization can reduce excessive dependency on specific neurons and improve model robustness (Amjadian, 2024; Karimi Dastgerdi & Zamani Boroujeni, 2020). The improved validation performance observed in the modified three-layer model further demonstrates the effectiveness of regularization and data balancing strategies in enhancing generalization capability.

The results of the present study are also important from a practical marketing and business perspective. Sentiment analysis systems capable of accurately processing Persian-language customer reviews can provide valuable insights into consumer preferences, product weaknesses, and market expectations in the Iranian automotive industry. Customer-generated online reviews represent a rich source of real-time information that can support product development, quality management, and customer relationship strategies. The findings therefore align with studies emphasizing the strategic role of sentiment analysis in brand management, customer satisfaction evaluation, and digital marketing analytics (Hossain, 2024; Selimovic et al., 2024). In addition, the use of deep learning-based sentiment analysis systems can assist automotive companies in monitoring public reactions to new products, evaluating consumer trust, and identifying competitive market opportunities.

The present findings also contribute to the growing literature on artificial intelligence applications in customer analytics and business intelligence. The integration of deep learning methods into sentiment analysis frameworks has transformed the ability of organizations to process unstructured textual data at large scale. Previous studies have highlighted the increasing role of AI-driven sentiment analysis in intelligent recommendation systems, digital marketing optimization, and predictive analytics (Mirzaee et al., 2024; Sudirjo et al., 2023). The present study extends this literature by demonstrating that optimized LSTM architectures can achieve highly accurate sentiment

classification in Persian-language automotive datasets despite linguistic complexity and data imbalance challenges.

Furthermore, the findings support the argument that deep learning-based sentiment analysis models remain highly relevant despite the emergence of transformer-based architectures such as BERT and ChatGPT. Although transformer models provide powerful contextual understanding capabilities, they often require extensive computational resources and large annotated corpora. The present study demonstrated that optimized LSTM architectures combined with preprocessing and augmentation techniques can still achieve highly competitive performance in domain-specific sentiment analysis tasks. These findings are aligned with previous research suggesting that LSTM networks remain efficient and effective solutions for sequential text classification, especially in low-resource language environments (Dam et al., 2025; Susnjak, 2024).

Another significant implication of this study concerns the development of Persian-language natural language processing systems. Most existing sentiment analysis studies have focused primarily on English-language datasets, while comparatively limited attention has been given to Persian-language deep learning applications. The present study therefore contributes to reducing this research gap by providing an optimized framework for sentiment analysis of Persian customer reviews in a specialized industrial domain. The findings are consistent with Iranian studies emphasizing the growing importance of social media analytics and text mining methods for understanding public attitudes and customer behavior in digital environments (Farajzadeh et al., 2023; Mohammadi, 2023).

One of the limitations of the present study was the restriction of the dataset to Iranian automotive platforms and Persian-language customer reviews. Although the dataset included diverse vehicle models and multiple online sources, the findings may not be fully generalizable to other industries, languages, or cultural contexts. In addition, the study relied primarily on textual comments and did not incorporate multimodal information such as images, emojis, audio content, or user interaction metadata. Another limitation concerns the imbalance of sentiment classes, particularly the relatively smaller representation of neutral comments, which may have affected classification performance despite the application of balancing techniques.

Future research may expand the scope of sentiment analysis by incorporating transformer-based architectures such as BERT, GPT-based systems, and hybrid deep

learning models for comparison with LSTM frameworks. Researchers may also investigate multilingual sentiment analysis systems capable of processing Persian and English customer reviews simultaneously. In addition, future studies could integrate multimodal data sources including images, voice comments, and behavioral interaction patterns to improve sentiment prediction accuracy. The use of larger datasets, real-time streaming data, and explainable artificial intelligence techniques may also enhance the interpretability and practical applicability of sentiment analysis systems in industrial environments.

From a practical perspective, the findings of this study suggest that automotive companies, marketers, and digital platform managers can employ LSTM-based sentiment analysis systems to monitor customer perceptions continuously and identify emerging market trends. These systems may support product quality improvement, after-sales service evaluation, customer complaint management, and targeted marketing strategies. In addition, policymakers and industry regulators may use sentiment analysis frameworks to evaluate public satisfaction with domestic automotive products and identify critical consumer concerns. The implementation of intelligent sentiment analysis systems can therefore contribute to improving customer-centered decision-making processes and enhancing competitiveness within the Iranian automotive industry.

Authors' Contributions

Authors contributed equally to this article.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethics Considerations

In this research, ethical standards including obtaining informed consent, ensuring privacy and confidentiality were considered.

References

- Amjadian, P. (2024). Remaining Useful Life Estimation of Bearings Using Vibration Signal Processing Based on Continuous Wavelet Transform and LSTM Deep Learning Network. *Transactions on Machine Intelligence*, 7(1), 61-69. <https://doi.org/10.47176/TMI.2024.61>
- Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A Comprehensive Survey on Sentiment Analysis: Approaches, Challenges and Trends. *Knowledge-Based Systems*, 226, 107134.
- Dam, M., Roy, S., Sarkar, D., & Mahata, S. K. (2025). Sentiment Analysis Using LSTM Model for Emotion Detection. Kolkata, India.
- Farajzadeh, K., Taghavifard, M. T., Toloei Ashlaghi, A., & Rashidi Komijan, A. (2023). Sentiment Analysis Based on Text Mining in the Social Network Twitter to Understand People's Thinking. *New Media Studies*, 9(34), 1-32. <https://doi.org/10.22054/nms.2023.69807.1453>
- Hossain, M. S. (2024). Textual Feature Engineering for Purchase Intent and Customer Satisfaction: Insights from Marketing 4.0 and Sentiment. *Sustainable Futures*, 8, 100385.
- Huang, B., Guo, R., Zhu, Y., Fang, Z., Zeng, G., Liu, J., Wang, Y., Fujita, H., & Shi, Z. (2022). Aspect-Level Sentiment Analysis with Aspect-Specific Context Position Information. *Knowledge-Based Systems*, 243, 108473.
- Karimi Dastgerdi, A., & Zamani Boroujeni, F. (2020). A Review of Deep Learning Methods for Financial Market Prediction. *Transactions on Data Analysis in Social Science*, 2(3), 164-172. <https://doi.org/10.47176/TDASS.2020.164>
- Kazemi, E., Sadeghzadeh, M., & Taeizadeh, A. (2023). Presenting a New Sentiment Analysis Method Based on Multi-Objective Archimedes Optimization Algorithm and Machine Learning. *Electronic and Cyber Defense*, 11(4), 145-153.
- Lin, Y., & Jen, L. (2024). The Impact of Consumer Heterogeneity in the Product Life Cycle on the Diffusion Patterns of User Reviews and Sales. *Journal of Retailing and Consumer Services*.
- Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment Analysis Methods, Applications, and Challenges: A Systematic Literature Review. *Journal of King Saud University - Computer and Information Sciences*, 36(4), 102048.
- Mehrkan, S. S., & Davari-Ardakani, H. (2025). An Integrated Model of Sentiment Analysis and Quantitative Index Data for Predicting Stock Market Trends: A Case Study of Tehran Stock Exchange. *Expert Systems with Applications*, 269, 126298.
- Mirzaee, A., Zeynali, M., Ghorbanzadeh, A., & Ghorbanzadeh, P. (2024). Personal Recommender Model and Predicting Consumer Behavior in Digital Marketing Based on Deep Learning. *Transactions on Machine Intelligence*, 7(3), 179-193. <https://doi.org/10.47176/TMI.2024.179>

- Mohammadi, M. (2023). Presenting a Predictive Model for Sentiment Analysis of Users toward the City Based on the Twitter Social Network: A Case Study of Iranian Metropolises. *Iranian Architecture and Urbanism*, 14(2), 301-317.
- Prasastio, F. R., Pradana, P. D. R., Permana, E., Afiahayati, & Frisky, A. Z. K. (2024). Sentiment Analysis of Electric Vehicle Incentives Using CNN-LSTM and SVM Models. Medan, Indonesia.
- Puh, K., & Bagic Babac, M. (2023). Predicting Sentiment and Rating of Tourist Reviews Using Machine Learning. *Journal of Hospitality and Tourism Insights*, 6(3), 1188-1204.
- Ricchiuti, F., & Sperli, G. (2025). An Advisor Neural Network Framework Using LSTM-Based Informative Stock Analysis. *Expert Systems with Applications*, 259, 125299.
- Rodriguez-Ibanez, M., Casanez-Ventura, A., Castejon-Mateos, F., & Cuenca-Jimenez, P. M. (2023). A Review on Sentiment Analysis from Social Media Platforms. *Expert Systems with Applications*, 119862.
- Selimovic, M., Abd Almisreb, A., & Amanzholova, S. (2024). The Role of Sentiment Analysis in Brand Management and Marketing: A Comparative Study. *Procedia Computer Science*, 251, 579-584.
- Sudirjo, F., Diantoro, K., Al-Gasawneh, J. A., Azzaakiyyah, H. K., & Ausat, A. M. A. (2023). Application of ChatGPT in Improving Customer Sentiment Analysis for Businesses. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5, 283-288.
- Susnjak, T. (2024). Applying BERT and ChatGPT for Sentiment Analysis of Lyme Disease in Scientific Literature. *Methods and Protocols*, 173-183.
- Wang, H., & Zheng, L. (2016). Sentiment Classification of Chinese Online Reviews: A Comparison of Factors Influencing Performances. *Enterprise Information Systems*, 10, 228-244.
- Wen, Z., Chen, Y., Liu, H., & Liang, Z. (2024). Text Mining Based Approach for Customer Sentiment and Product Competitiveness Using Composite Online Review Data. *Journal of Theoretical and Applied Electronic Commerce Research*, 19, 1776-1792.
<https://doi.org/10.3390/jtaer19030087>