Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 4 (2024)

A Survey of Sentiment Analysis Using Various Machine Learning and Deep Learning Techniques

Dr. Chitra K.[1*], Shameema Rahmath K. T.[2]

[1*]*Professor & HOD, Department of Computer Science, Sri Krishna Adithya College of Arts and Science, Kovaipudur, Coimbatore, India*

[2]*Reasearch Scholar, Department of Computer Science, Sri Krishna Adithya College of Arts and Science, Kovaipudur, Coimbatore, India*

Abstract

In the modern era, Deep Learning (DL) and Machine Learning (ML) emerged as the advanced strategies for applying various Natural Language Processing (NLP) to enhance the performance of sentiment analysis. Main of this sentiment analysis is to extract the essential opinions as well as attitudes towards an entity. This process is one of the most powerful tool utilized by marketing, governments, business and others. Also, traditional sentiment analysis approaches are mainly focused on text content. Nevertheless, advance technology have permitted to prompt their feelings and expressions through video, audio, texts and images. As the most crucial task for improving decision-making, sentiment analysis involves identifying the underlying sentiment or opinion of data. Though sentiment analysis has advanced recently, there are still issues with current models, including excessive dimensionality, negation handling, domain dependence, and ineffective keyword extraction. This study looks at several viewpoints about the development and application of a successful sentiment analysis model such as sequential ML and DL techniques and conducts the detailed review.

Keywords: *Machine Learning, Deep Learning, Natural Language Processing, Methodology*.

1. Introduction

Sentiment analysis (SA), often known as opinion mining, is a field of study that examines people's feelings or perspectives about various entities, including subjects, people, events, problems, services, goods, organisations, and their characteristics. Sentiment analysis has become increasingly important due to the expansion of information available on social media; relevant research has been categorised into three primary areas of application.
From a business standpoint, sentiment analysis can offer online suggestions and counsel to consumers as well as retailers. On the one hand, e-commerce platforms can use the consumer preferences revealed by the data to analyse their offerings.

However, internet purchasing is virtual, it might be challenging to fully and impartially comprehend a product and determine whether a customer is prepared to read through the reviews and opinions of other customers. From a political stance, the enormous need for political data might be considered to be a further crucial element. People search for or express opinions online for reasons other than commercial gain. Sentiment analysis is a way to identifying ambiguity in words, opinions, etc. Sentiment analysis provides information about a representative's and user's opinions on a certain topic. A writer's expressive style or tone choice conveys their thoughts and emotions. A plethora of algorithms has recently been invented to analyse, forecast, and assess sentiment from textual data, including customer or product reviews. Sentiment analysis has the potential to greatly facilitate the process of polarity recognition. It also has issues with domain dependency, negation, bipolar terminology, the weight of NLP, spam and misleading data, and a large vocabulary. In order to organise the text and extract information that text-mining heuristics and machine learning algorithms may employ later, sentiment analysis requires preprocessing elements. Remaining section of this paper is organized as follows, section.2 provides various DL and ML frameworks, methodology summary is discusses in section.3 and section 4 provides the discussion of the outcome parameters. Finally, the paper is ends with results and conclusion section.5.

2 Literature survey

In the literature, numerous ML, DL and metaheuristic strategies have been developed

In the modern era, numerous people use social media platform, so offensive contents may be increased. Consequently, classification of offensive content is the most challenging task towards the sentiment analysis system. Therefore, to classify the offensive text from the content a novel Long Short Term based BOOST (LST-BOOST) network is developed by Wadud et al. [6]. Moreover, AdaBoost classifier with Principle Component Assessment (PCA) is adapted with this network to attain the most important variance and decrease the weighted error. Additionally, several word embedding baseline strategies are used for comparison process. Here it is attains 92.61% of f-1score.

Managing the big datasets and various textual data are challenging task in sentiment analysis. To enhance the handling accuracy and efficiency of sentiment analysis process Smitha et al, [7] have developed the Modified Bayesian with Weight Guided boosting algorithm. Here, the boosting strategy is used to manage the lengthy sentence and network is utilized as to reduce the challenges facing the sentiment analysis. Moreover, Weight Guided approach is used to detect the discriminative and relevant features from the text. Consequently, the proposed technique is to optimize and enhance sentiment analysis performance.

Serpil Aslan et al, [8] have proposed a Convolution neural based Arithmetic Optimization algorithm was used extract the features like long sentence from the review. Initially, open access covid-19 datasets were collected and enabled as the pre-processing process. Then, fastText strategy is used as the dimensional feature representation. At last classification was compared with existing SA approaches to prove the highest performance.

In today’s world, network system are used most interesting communication technology as social media network like Facebook, twitter, etc. As a results, various useful information is acquired from these networks. From that opinion classification is the important to express the people suggestion. Therefore, Meena et al, [9] Convolution Neural Network (CNN) were used to perform the SA process efficiently. Here, TV, Laptop and Phone review datasets were collected and perform he different applications.

Analysing the customer review on social media networking is one of the important part to improve the business application. To analyse customer review Behera et al, [10] hybridization of CNN with LSTM framework suggested for SA process. Here, CNN model is used as feature selection and LSTM is applied for two objectives such as scalability and adaptable.

To address the multi text and binary classification problems Sanskar soni et al, [11] have developed the text based CNN (TextConvoNet) to provide the better outcomes compared to other techniques. Here, one dimensional filters are used to extract the n-gram features and sentence matrix are used to capture the text data. Consequently, it utilizes two dimensional convolution function to perform as alternative approach. Also, the proposed model is further expanded by combining ML, DL and attention models.

In a daily life cycle human being has continuously uses the social media communication networks to express their opinions, feedbacks and feelings. In this, sentiment analysis is important role to determine the text classification such as positive, negative and neutral. Therefore, Kian long tan et al, [12] have suggested the optimized LSTM based Bidirectional Encoder Representation from Transformers (BERT) approach to control the imbalanced dataset issues. Moreover this hybrid is the combination of transformer model and sequence model.

Aksh patel et al, [13] have introduced a BERT is utilized to attain the pre-trained bidirectional representation function for combining the input text. Here, right and left side tokens are quickly fine-tuned and tokenized due to the training phase. The proposed framework has classified into two phases such as BERT base as well as BERT large. Both are adapted in three different layers such parameter layer, transformer layer and attention layer.

To enhance the marketing field based on the product review Sangeetha et al, [14] have introduced a LSTM based Taylor Harris Hawks optimization strategy (LSTM-THHO). Here, optimization algorithm is combined with memory network to improve the classification performance. Moreover, hidden layers weighting factors are enhanced to reduce the nose features from the collected dataset. It is having three basic steps such as pre-processing, optimal selection using and termination hybrid techniques.

In the data mining, sentiment analysis is the trendiest research topic so, Nikhat et al, [15] have developed the gated attention frameworks to provide effective and efficient feature selection model. Here, 140 datasets were used to fit into the initial layer and performed pre-processing function. After that, LSTM based modified inverse class frequency strategy is utilized to extract the features from the pre-processed data. Then, feature selection process is done with the help of hybridization of white shark with mutation optimizer.

Myagmarsuren et al, [16], have introduced a convolution with LSTM and Ant Lion Optimization strategy is used to improve the classification of English text. Here, loop core network is acquires the typical text for the English text during the training stages. Moreover, two types of experiments were done such as dependability and viability to provide the relevant features. Consequently, experimental outcomes were analysed the source data and statistical calculation provides similar public data.

Sayar Singh shekhawat et al, [17] hybridized K-means clustering based spider monkey optimization approach is to achieve the finest cluster head selection on dataset. Here two types of dataset were used such as twitter and sender2. Moreover, the clustering based metaheuristic approach efficiently solve the data mining issues.

To analyse the stock market data, Nan Jing et al, [18] have developed the CNN based LSTM approach can classify the sentiment analysis parameters from the stock market data. Moreover, Shanghai Stock Exchanges (SSE) intervals are used to validate the applicability and effectiveness of the developed model. Then, the experiment results of the proposed classifier is compared with conventional classifier. From the comparison the proposed classifier has attained finest outcomes.

In a NLP, sentiment classification is the most important investigation area therefore, Ishaani priyadarshini et al, [19] have developed the CNN with LSTM based Grid Search (GS) DL framework is to understand the customer emotions, opinions and attitudes. Here, hyperparameters optimization approach to reduce the losses and improve the accuracy. Hence, GS based, hyperparameters optimization led to enhance the overall efficiency. Proposed study has report optimum accuracy for all collected dataset with respect to existing baseline strategies.

Hichem Rahab et al, [20] have proposed a rule based binary equilibrium with metaheuristic optimization strategy were created and performed the sentiment analysis process on Arabic document. Here, the newly developed algorithm was efficiently perform the existing research gap such as lower classification and low resourced languages. Here 80% of data used as training and 20% of data used as testing. Figure 1 depicts the classification of ML and DL models.

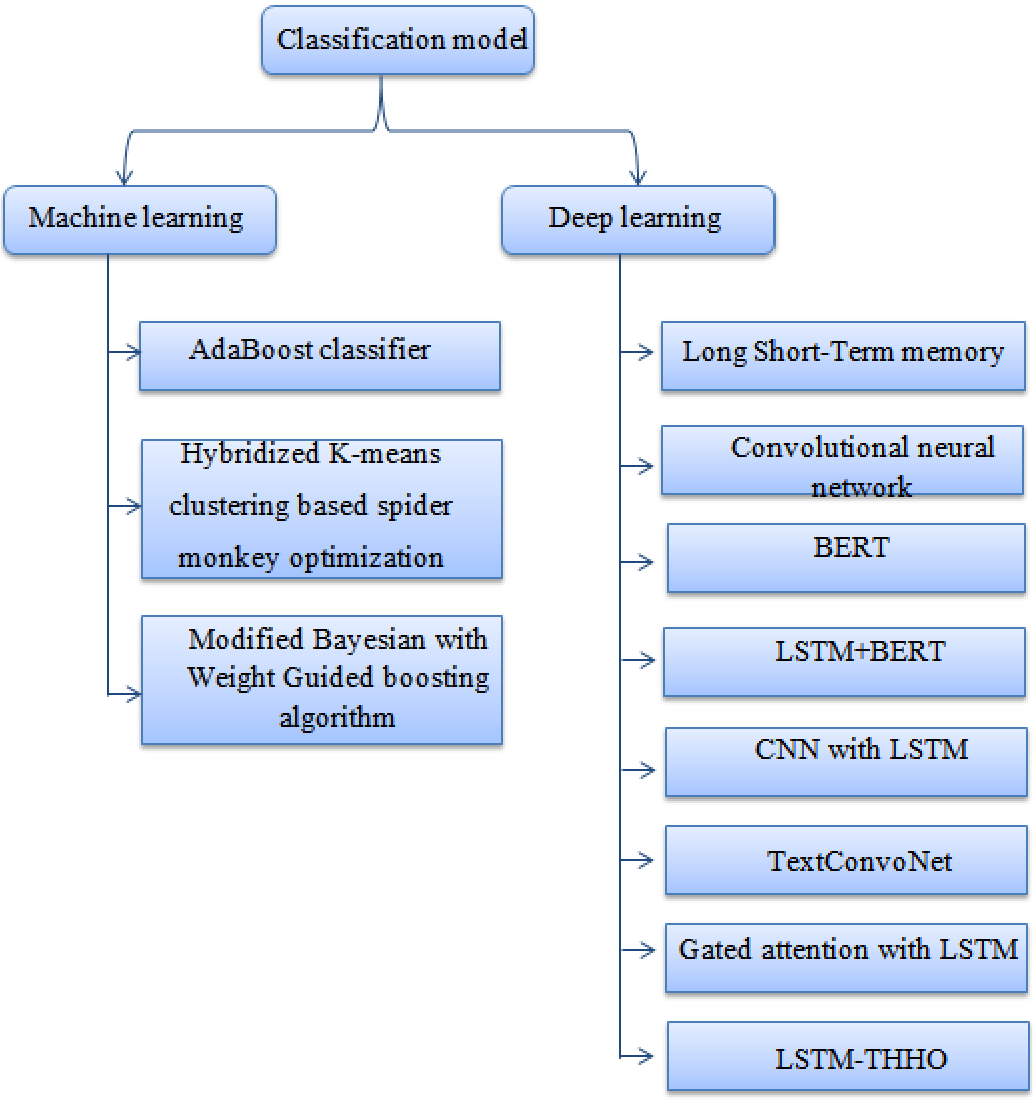

Figure 1: Classification of ML and DL models

Table.1 Summary of ML techniques used for sentiment analysis

Sl.no	**Author name**	**Year**	**Proposed technique**	**Advantages**	**Limitations**	**Findings (%) and dataset**
1	Smitha et al, [7]	2023	Modified Bayesian with Weight Guided boosting algorithm	Enhanced efficiency, scalability and accuracy of sentiment analysis	It is failed to handle the imbalance data and cannot provide promising results	Accuracy:98.49
2	Sayar Singh shekhawat et al, [17]	2021	hybridized K-means clustering based spider monkey	Better elimination of unnecessary information,	Failed to observe the language context	Accuracy:88.93

			optimization approach	enhanced the accurate level of acronyms as well as stop words		

Table.2 Summary of DL techniques used for sentiment analysis

Sl.no	**Author name**	**Year**	**Proposed technique**	**Advantages**	**Limitations**	**Findings (%) and dataset**
1	Serpil Aslan et al, [8]	2023	Convolution neural based Arithmetic Optimization algorithm	Effective sentiment analysis, accurately detect the finical data opinions	Complex to update the dictionary, not applicable for all sentiment classification domains	Accuracy:93.5 Precision:90.2 Recall:90.1 F-measure:90
2	Meena et al, [9]	2022	CNN	It provides better classification results and intermediate opinion	Restricted accurate level when datasets size increased	Accuracy:95%
3	Sanskar soni et al, [11]	2023	TextConvoNet	Enhanced the performance, lower error rate, enhanced binary classes.	This type of networks has introduced many computational complexity	Accuracy:92.9 Precision:93 Recall:93 F-measure:93
4	Aksh patel et al, [13]	2023	BERT	The proposed work provides higher results in terms of accuracy, precision, recall and F-measure	For their proposed algorithm consideration has taken more investigation study during the training phase	Accuracy:83 Precision:84 Recall:90 F-measure:85

Table.3 Summary of ensemble techniques used for sentiment analysis

Sl.no	**Author name**	**Year**	**Proposed technique**	**Advantages**	**Limitations**	**Findings (%) and dataset**

1	Wadud et al. [6].	2022	Long Short Term based BOOST (LST-BOOT) and AdaBoost classifier with Principle Component Assessment (PCA)	This method is used to observe the numerous messages, comments and posts from the social media platforms	It is failed to manage the infinitely longer sentence in the text	Accuracy:92.11 Precision:89.48 Recall:93.38% F-measure:91.395
2	Kian long tan et al, [12]	2022	LSTM with BERT	It is utilized to create the extra lexically same samples, remove irrelevant words from the corpus	Remove the details of particular aspect	Accuracy:89.7 Precision:90 Recall: 90 F-measure: 90
3	Sangeetha et al, [14]	2023	LSTM-THHO	More sensitivity and higher specificity, effective classification outcomes, lower validation loss	Large number of datasets cannot applicable for this model	Accuracy:96.58 Precision:95.6 Recall: 95.6 F-measure: 95.6
4	Behera et al, [10]	2021	hybridization of CNN with LSTM framework	Highly adaptable for longer social media data, easy to manage, manage all types of dependencies	Convolution layer fails to observe the sequence dependence words from the text	Accuracy: 90% F-1 score:81% specificity:92%
5	Nikhat et al, [15]	2023	gated attention frameworks and LSTM based modified inverse class frequency strategy hybridization of white shark with mutation optimizer	Better performance compared to other classifiers such as CNN, RNN	In some cases, it fails to provide essential features and his can be achieved loss, decreased the effective performance,	Accuracy:97.86 Precision:96.65 Recall:96.76 F-measure:96.70

6	Myagmarsuren et al, [16]	2023	convolution with LSTM and Ant Lion Optimization strategy	Enhanced text classification results, quickly retrieve the accurate information, easy to handle	Lot of training data can leads to cause the training time	Accuracy: 90 Precision: 92 Recall: 88 F-measure:92
7	Nan Jing et al, [18]	2021	CNN based LSTM approach	More accurate prediction,	Further improves the prediction power in future	Accuracy:89 f-measure:84
8	Ishaani priyadarshini et al, [19]	2021	CNN with LSTM based Grid Search (GS) DL framework	Enhanced performance such as grammatical variations, misspellings, complicated structures and slang	Hyperparameters tuning is poor performance	Performance evaluation for dataset 1 and 2; Accuracy:96.4 and 97.8 Precision:98.95 and 98.2 Sensitivity:97.4 and 98.9 Specificity:99.2 and 99 F-measure:98.1 and 97.2
9	Hichem Rahab et al, [20]	2023	rule based binary equilibrium with metaheuristic optimization strategy	Effective sentiment analysis, efficiently remove the redundant as well as irrelevant features	Huge amount of attributes can reduce the size of the classifier and take too much time	Accuracy:84 Recall:68 precision:68 Classifier size:13

3. Methodology

According to traditional research, the most common use of ML and DL approaches is sentiment analysis for social media. Traditional ML techniques such as Naïve Bayes, KNN are used in different applications. Ensemble classifiers are employed in addition to traditional machine learning techniques to obtain more accurate results. Furthermore, DL approaches have been receiving more and more attention. But DL techniques are not perfect in two important areas. One is that CNN algorithms perform differently depending on the volume of annotated examples. The other is their reliance on a certain domain. For this reason, LSTM networks are used since they don't need annotated data. However, these methods rely on static word lists, and dictionary usage is context-independent. Hybrid techniques such as LST BOOST with AdaBoost classifier, CNN with LSTM, LSTM-THHO etc are thus proposed to compensate for the shortcomings of LSTM networks and optimization approaches. Furthermore, the social network and its attributes are proposed to be utilized through graph-based approaches. These methods do not require large amounts of annotated data. However, the emotion lexicons and the connection graphs are domain specific, they are dependent on the domain.

4. Discussion

The parameters utilized for evaluating the performance of the sentiment analysis framework are discussed in this section. Currently, the ML and DL methods are employed to categorize the sentiments or opinions on the social media platforms. The classification performances of these methods are assessed in terms of accuracy, precision, recall or sensitivity, f-measure, specificity, and false positive rate (FPR). The definition for these parameters are described below,

Accuracy: It measures the overall correctness of the sentiment classification. It quantifies how effectively the classifier recognizes the sentiments in social media platforms, and it is determined as the ratio of correctly identified instances to the total instances.

Precision: It defines the efficiency of the classifier in accurately predicting the positive cases (true positives). It indicates the ratio of true positives to the sum of true and false positives.

Recall (Sensitivity): It measures the model’s capacity to capture all relevant instances for sentiment classification. It is calculated as the ratio of true positive predictions to the total actual positive instances.

F-Measure: F-measure or F1-score represents the harmonic mean of recall and precision. It delivers a balanced analysis of the model's outcome considering both positive and negative cases.

Specificity: It measures the model’s capacity to predict the negative instances accurately. It is defined as the proportion of true negative predictions made by the system to the total actual negative instances.

FPR: It measures the percentage of incorrect positive predictions made by the classifier among real negatives, and it is defined as the proportion of false positive predictions to the total negative instances.

The evaluation of these parameters enables us to assess the model’s performance in analyzing the sentiments on social media platforms. LSTM approaches has attained 92.115 accuracy, 89.48% of precision, 93.38% recall and 91.395% of f-measure. But, CNN with optimization approaches has attaining better performance that the other. Moreover, single CNN model has achieved 95% accuracy and single BERT model has accuracy 83%, precision 84%, Recall 90% and 85% f-measure, which is compared to CNN model.

5. Conclusion

Sentiments or opinion classification performance can provide most inclusive information from the social media platforms. ML and DL techniques have its own limitations, hence a hybrid approach is proposed that combine ML and DL techniques. The hybrid approach incorporates domain knowledge and feature engineering and offer promising avenues to address the limitations of individual methods and enhance sentiment analysis performance. The proposed hybrid technique uses Convolutional Neural Network (CNN) in combination with other standalone classifiers such as an AdaBoost classifier, is an interesting approach that can potentially yield good results.

Compliance with Ethical Standards

Conflict of interest

The authors declare that they have no conflict of interest.

Human and Animal Rights

This article does not contain any studies with human or animal subjects performed by any of the authors.

Informed Consent

Informed consent does not apply as this was a retrospective review with no identifying patient information.

Funding: Not applicable

Conflicts of interest Statement: Not applicable

Consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and material:

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Code availability: Not applicable

Reference

1. Cui, Jingfeng, et al. "Survey on sentiment analysis: evolution of research methods and topics." Artificial Intelligence Review (2023): 1-42.
2. Chan, Jireh Yi-Le, et al. "State of the art: a review of sentiment analysis based on sequential transfer learning." Artificial Intelligence Review 56.1 (2023): 749-780.
3. Gandhi, Ankita, et al. "Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions." Information Fusion 91 (2023): 424-444.
4. Zhu, Linan, et al. "Multimodal sentiment analysis based on fusion methods: A survey." Information Fusion 95 (2023): 306-325.
5. Mercha, El Mahdi, and Houda Benbrahim. "Machine learning and deep learning for sentiment analysis across languages: A survey." Neurocomputing 531 (2023): 195-216.
6. Wadud, Md Anwar Hussen, et al. "How can we manage offensive text in social media-a text classification approach using LSTM-BOOST." International Journal of Information Management Data Insights 2.2 (2022): 100095.
7. Nayak, Smitha, and Yogesh Kumar Sharma. "A modified Bayesian boosting algorithm with weight-guided optimal feature selection for sentiment analysis." Decision Analytics Journal 8 (2023): 100289.
8. Aslan, S., Kızıloluk, S. & Sert, E. TSA-CNN-AOA: Twitter sentiment analysis using CNN optimized via arithmetic optimization algorithm. Neural Comput & Applic **35**, 10311–10328 (2023). https://doi.org/10.1007/s00521-023-08236-2
9. Meena, Gaurav, Krishna Kumar Mohbey, and Ajay Indian. "Categorizing sentiment polarities in social networks data using convolutional neural network." SN Computer Science 3.2 (2022): 116.
10. Behera, Ranjan Kumar, et al. "Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data." Information Processing & Management 58.1 (2021): 102435.
11. Soni, S., Chouhan, S.S. & Rathore, S.S. TextConvoNet: a convolutional neural network based architecture for text classification. Appl Intell **53**, 14249–14268 (2023). https://doi.org/10.1007/s10489-022-04221-9
12. Tan, Kian Long, et al. "RoBERTa-LSTM: a hybrid model for sentiment analysis with transformer and recurrent neural network." IEEE Access 10 (2022): 21517-21525.
13. Patel, Aksh, Parita Oza, and Smita Agrawal. "Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model." Procedia Computer Science 218 (2023): 2459-2467.
14. Sangeetha, J., and U. Kumaran. "A hybrid optimization algorithm using BiLSTM structure for sentiment analysis." Measurement: Sensors 25 (2023): 100619.
15. Parveen, N., Chakrabarti, P., Hung, B.T. et al. Twitter sentiment analysis using hybrid gated attention recurrent network. J Big Data **10**, 50 (2023). https://doi.org/10.1186/s40537-023-00726-3

16. Orosoo, Myagmarsuren, et al. "Performance analysis of a novel hybrid deep learning approach in classification of quality-related English text." Measurement: Sensors (2023): 100852.

17. Shekhawat, S.S., Shringi, S. & Sharma, H. Twitter sentiment analysis using hybrid Spider Monkey optimization method. Evol. Intel. **14**, 1307–1316 (2021). https://doi.org/10.1007/s12065-019-00334-2

18. Jing, Nan, Zhao Wu, and Hefei Wang. "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction." Expert Systems with Applications 178 (2021): 115019.

19. Priyadarshini, I., Cotton, C. A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis. J Supercomput **77**, 13911–13932 (2021). https://doi.org/10.1007/s11227-021-03838-w

20. Rahab, H., Haouassi, H. & Laouid, A. Rule-Based Arabic Sentiment Analysis using Binary Equilibrium Optimization Algorithm. Arab J Sci Eng **48**, 2359–2374 (2023). https://doi.org/10.1007/s13369-022-07198-2