Chapter 4 Machine learning for sentiment analysis in social networks data: Advances and perspectives

Capítulo 4 Aprendizaje automático para el análisis de sentimiento en datos de redes sociales: Avances y perspectivas

PÉREZ-PÉREZ, Anabel†[1], CALVARIO-SÁNCHEZ, Gabriela *[1] and ALARCÓN-MARTÍNEZ, Teresa E[2].

[1]*Department of Electronics, Systems, and Informatics, ITESO—The Jesuit University of Guadalajara, Tlaquepaque, Jalisco 45604, México.*
[2]*Department of Computer Science and Engineering, The Valleys' University Center, Ameca, Jalisco 46600, México.*

ID 1st Author: *Anabel, Pérez-Pérez* / **ORC ID**: 0000-0002-0431-5136, **CVU CONAHCYT ID:** 1137076

ID 1st Co-author: *Gabriela, Calvario-Sánchez* / **ORC ID**: 0000-0003-2864-4839, **CVU CONAHCYT ID:** 381923

ID 2nd Co-author: *Teresa, E. Alarcón-Martínez* / **ORC ID**: 0000-0002-7399-9068, **CVU CONAHCYT ID:** 249567

DOI: 10.35429/H.2023.8.55.68

A. Pérez, G. Calvario and T. Alarcón

*gabriela.calvario@iteso.mx

A. Marroquín, L. Castillo, J. Olivare sand G.Morado (Coord) Engineering Sciences and Applications. Handbooks-

Abstract

Sentiment analysis is a field of study within artificial intelligence aimed at comprehending opinions and emotions expressed in natural language, such as texts published on social networks. Social networks are understood as online technologies, tools, and applications that allow users to generate content, share and exchange information, and create interpersonal and communal relationships through the Internet. The data generated from these sources are highly intricate to analyze, hence the relevance of computational tools. Multiple approaches exist that tackle sentiment analysis through artificial intelligence, specifically through machine learning. In this chapter, a literature review and state-of-the-art analysis were conducted regarding sentiment analysis on social network data, with the objective of identifying technologies that exhibit superior performance in this task; the available methodologies are cited to facilitate the selection of an appropriate method, and the advantages and disadvantages of all reviewed methodologies are enumerated. The guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology were applied and finally, the systematization of knowledge was carried out on 21 articles.

Sentiment analysis, Artificial intelligence, Machine learning, Data from social networks, Text processing

Resumen

El análisis de sentimientos es un área de estudio de la inteligencia artificial para comprender opiniones y emociones expresadas en lenguaje natural, como textos publicados en redes sociales. Las redes sociales se entienden como tecnologías, herramientas y aplicaciones en línea que permiten a los usuarios, a través de Internet, generar contenidos, compartir, intercambiar información, crear relaciones interpersonales y comunitarias. Estos datos generados son muy complejos de analizar, por lo que es pertinente el uso de herramientas computacionales. Existen varios enfoques que abordan el análisis de sentimientos a través de la inteligencia artificial y específicamente el aprendizaje automático. En este capítulo se desarrolló una revisión de la literatura y estado del arte sobre análisis de sentimiento en datos de redes sociales, con el fin de identificar tecnologías con el mejor desempeño en el desarrollo de esta tarea; se citan las metodologías disponibles, para facilitar la selección del método apropiado y se enuncian las ventajas y desventajas de todas las metodologías revisadas. Se aplicaron los lineamientos de la metodología Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) y finalmente se realizó la sistematización del conocimiento sobre 21 artículos elegidos.

Análisis de sentimientos, Inteligencia artificial, Aprendizaje automático, Datos de redes sociales, Procesamiento de texto

1. Introduction

Sentiment analysis (SA) refers to the application of Natural Language Processing (NLP), computational linguistics and text analytics to identify and extract subjective information in source materials, like opinions from the text and classify the polarity of subjects into positive, negative, or neutral to determine the public group perception. NLP is a field of Computer science, Artificial intelligence (AI) and Linguistics concerned with the interactions between computers and human natural languages. Extracting useful knowledge from naturally written texts, allows NLP to resolve the distance between humanity and machine. The goal of SA is to extract and analyze knowledge from personal data or reviews and feedback provided from different sources of data on the internet (Batrinca & Treleaven, 2015), (Xue et al., 2023), (Yadav, 2023).

SA is divided into three different levels which are sentence level, document level and feature level. The purpose is to classify the opinion either from sentence, document or features into positive and negative sentiment or even neutral. Therefore, the methods are divided into lexicon-based, machine learning-based, hybrid methods and, more recently, deep learning-based approaches. Machine learning-based approach utilized algorithms to extract and detect sentiment from a data, while lexicon-based approach works by counting the positive and negative words that relate to the data and uses a glossary of sentiment terms including enhancement and negation to measure the polarity of each phrase.

However, this method depends on the extraction of knowledge from a statement with an opinion polarization. The deep learning-based approach has two phases, the term embedding in the text corpus is learned in the first phase and the second phase focuses on the use of word embedding to create interpretations of sentences of semantic composition using different deep learning techniques (Drus & Khalid, 2019), (Sharma et al., 2019), (AlBadani et al., 2022), (Xue et al., 2023).

SA is improved, considerably, by the application of Machine Learning (ML), with supervised learning and unsupervised learning methods; where a computer program is assigned to perform some tasks and it is said that the machine has learnt from its experience if its measurable performance in these tasks improves as it gains more and more experience in executing these tasks. So, the machine makes decisions and does predictions / forecasting based on data. The supervised learning approach is used when there is labeled data available for training the model and the unsupervised learning method is used when the reliability of labeled data is difficult (Sharma et al., 2019), (Ray, 2019), (Yadav, 2023).

At present, social media data is the largest, richest and most dynamic evidence base of human behavior, bringing new opportunities to understand individuals, groups and society. The use of social networks generates massive data characterized computationally by big size, noise, and dynamism; now most frequently named as big social data. These characteristics make it very complex to analyze, resulting in the pertinent use of computational means to it (Batrinca & Treleaven, 2015) (Xue et al., 2023).

The purpose of SA on data extracted from social networks is to recognize potential drift in society as it concerns the attitudes, observations, and the expectations of the populace (Batrinca & Treleaven, 2015). SA with ML techniques on social network data is increasingly used by organizations to understand the opinion and emotions of users expressed on social networks and to understand where the organization is headed and when the institution needs to change strategies to be more efficient and effective (Carvalho & Plastino, 2021), (Shofiya & Abidi, 2021). Its application ranges from monitoring a brand's reputation to monitoring public opinion on political or social issues; to the identification of patterns and trends in user opinions, which can be useful for decision-making (Ansari & Khan, 2021), (Chandrasekaran et al., 2021), (Shekhawat et al., 2021), (Xue et al., 2023), (Yadav, 2023).

The objective of this study is to analyze the empirical and practical evidence available and provide a better understanding of the application of ML for SA expressed in social networks, by examining related literature published between 2018 and 2022; to determine the scientificity, validity of its use and social impact in other contexts, which will allow determining the state of the art based on the behavior of said phenomenon. This study does not intend to define concepts or substantiate the theoretical aspects of the approaches, models or algorithms summarized in the reviewed research. This manuscript is intended for researchers in computer science and information technology, who are assumed to be familiar with the terms discussed here.

The main contributions are the literature review, in order to identify technologies with the best performance to the SA in data from social networks; the available methodologies are cited to facilitate the selection of the appropriate method for a specific sentiment analysis task, giving advantages and disadvantages of all reviewed methodologies. The text is organized as follows: Section 2 describes the methodology, Section 3 details about the different approaches for sentiment analysis, Section 4 summarizes the challenges related to sentiment analysis, and finally in Section 5 the conclusions are given.

2. Methodology

This study was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement and aimed to conduct a meta-analysis for assessing a systematic review of published scientific literature on ML for SA expressed in social networks, between 2013 and 2022, focusing attention on the last five years of the mentioned period (Page, 2021). The initial search for published research on the subject was carried out in December 2022, in Google Scholar Web Page, with the purpose of determining the use of the terms and their combination. The terms used initially were 'sentiment analysis' and 'social media' and later, it was extended with a combination, using the boolean operators AND and OR as appropriate, of the terms 'opinion mining', 'social network', 'artificial intelligence'. These searches offered a global vision of the breadth of the subject.

In March 2023, the search was bounded to publications from the last five years, from 2018 to 2022 inclusive; and the results found in databases such as the IEEE, Scopus, ScienceDirect and Web of Science were added, with the purpose of covering the largest number of articles that should have been included. The combination of terms that yielded the best results was the following: ((sentiment analysis OR opinion mining) AND (social media OR social network)) AND (artificial intelligence OR machine learning). Finally, 678 results were obtained in this search.

The searches in SciSpace and Elicit were also considered. Both sites use AI algorithms for facilitated searching and generate a general and specific summary of the most relevant components and contributions of the resulting articles and research, as well as essential data of each of them. According to the website itself, Elicit is a research assistant using language models like GPT-3 to automate parts of researchers' workflows (Ought, 2023).

If you ask a question, Elicit will show relevant papers and summaries of key information about those papers in an easy-to-use table, whereas SciSpace is the easiest way to find, understand, and learn any research paper (SciSpace, 2023). In general, these applications improve the selection of the articles for the review and the management of references. For every paper, it gets simple explanations and answers using AI and discovers a network of connected and relevant papers (Kung, 2023).

Through Elicit and SciSpace, 125 results were obtained and limited in dates. Considering the non-veracity or authenticity of these tools by the scientific community, the results were only considered for the triangulation of the information and the initial readings of the abstracts to support the defined inclusion and exclusion criteria. The integration of AI in these sites is considered a contribution in the development of the systematic review, undoubtedly beneficial for the scientific community. During the entire review process, the total number of articles processed was 803, based on the defined inclusion and exclusion criteria.

The inclusion criteria defined for the selection of articles include: AI techniques and mainly ML used to predict sentiment behavior in text-type data; published between 2018 and 2022, including empirical researches and not reviews, books or manuals, reports, methodologies or tools for sentiment analysis with data obtained from social networks.

Among the exclusion criteria are: duplicated works; papers in which ML and AI are not used, publications without open access; studies limited to proposals; conceptual reviews and theoretical foundations papers without strategies and methodologies or activities for sentiment analysis; works in which data are not obtained from social networks; papers focused on industrial, manufacturing, or production of tangible goods sectors; research that does not draw conclusions on the behavior of the sentiment analysis.

According to these criteria, initially, 407 articles were considered adequate. Subsequently we eliminated 158 duplicates; 185 by reading the title and 53 for not being able to access the full text of the publication. The remaining articles were studied, and from the reading, 33 articles were discarded for being limited proposals, conceptual reviews and theoretical foundations; 119 for not applying data obtained from social networks; and 29 for not using ML for data processing.

In a more detailed, but fragmentary or partial reading, 63 were discarded for not specifying the methodology to be used; 111 for being studies focused on industrial, manufacturing, or production of tangible goods sectors; and 31 because they are only texts that justify the need for analysis and do not yield conclusions on the behavior of feelings or the accuracy of the methods used.

Nineteen articles satisfied the inclusion criteria, indicating the use of ML, as a technique for processing feelings expressed in social networks. Taking into account these nineteen articles we carry out the detailed reading necessary for the development of the systematic review. After reading the selected 19 articles in depth, based on their references, 2 new articles were included, considering them important for the review; therefore, in total, the systematization of knowledge was carried out on 21 articles.

3. Results

The final articles, chosen for the review, were separated for review and summary, considering the objectives pursued by the authors, which allowed a better understanding of the results. Three of the investigations presented a study of topics related to sentiment analysis and the algorithms, methods and approaches used. Two of them conceptually addressed the fundamentals, provided a global vision of the sentimental analysis task and compared the results shown by different authors in terms of techniques and methods of sentiments according to the specified contexts and data. The periods studied by these authors range from 2013 to 2020, inclusive.

Consequently, most of the investigations of the articles studied were relevant and interesting. The results presented in these studies serve as a basis to strengthen the theoretical and practical knowledge related to the subject. The main conclusions drawn from the investigations, reflected in the articles studied, are summarized in Table 1.

In Sharma et al.'s study (2020), the authors discussed the methods for sentiment classification and gave a comparison of algorithms experimented by different researchers on different datasets along with performance measures. As a result of the research, the authors concluded that the "term presence" is more important than the "term frequency" in SA; adjective, adverb, and verbs can be considered as features and irrelevant words can be removed from the corpus so that vocab size can be reduced (Mejova & Srinivasan, 2011). In addition, the authors indicated that most researchers performed SA using English language; but they found some researchers used non-English languages for solving SA problems providing compatible results as well (Che et al., 2015; Sharma et al., 2015; Sumit et al., 2018, as cited in Sharma et al., 2020).

Meanwhile, the perspective of Babu and Kanaga (2022), is also interesting. They addresses the impact of feature and detail extraction on the techniques used in the classification stage of SA; and the handling of emoticons and emojis in the text, referring to the fact that NLTK Tokenizer should be used to tokenize data from social networks into individual words where all emoticons and emojis are saved without deleting them (Li et al., 2023). These authors studied thirteen articles, in which ML techniques are utilized for the analysis of sentiments in data obtained from social networks. Among the techniques applied on Facebook datasets: Support Vector Machine (SVM), Naive Bayes (NB), K- Nearest Neighbors (KNN), and Random Forest (RF). In that case, the method with the best performance was RF with an accuracy of 84.6%. On Twitter datasets, ten studies in the main article applied SVM, five applied NB, four applied RF and four applied Multimodal NB. The highest results, on that datasets, were for applying SVM with 93.1% accuracy and Multimodal NB with 92.2% accuracy.

Previous results are reaffirmed in the investigation of Drus and Khalid (2019), in which it is pointed, that the most used method in ML application for SA, is the SVM and NB model. They explained that NB is successful when applied on well-formed text corpus, while SVM gives a good performance for low shape dataset. Moreover, it is explicit in their review that most of the study adapted SentiWordnet and Term Frequency-Inverse Document Frequency (TF-IDF) methods, when conducting sentiment analysis (Agarwal et al., 2015), (Rahman et al., 2022). Nonetheless, they affirm that, ML method performs poorly on Facebook with people posting in random length and lots of spelling mistakes and it requires a huge amount of training samples to adapt the method as the amount of dataset will influence the size and quality of the output.

The remaining reviewed articles were selected, after meeting the inclusion criteria, for presenting diverse and notable methodologies among the results published in the reviewed period. The performance value of the studied ML methods for SA on different dataset is summarized in Table 2. The reported results make evident that the authors agree that sentiments can be extracted, processed and classified, using a wide variety of algorithms and they recognize that the lexicon-based approach, standard machine-based approach, and deep learning-based approach are the three main approaches in SA. In most of the publications summarized in Table 2, a comparison is made between algorithms, methods and models to be applied for SA; highlighting, in this sense, the application of Artificial Neural Networks (ANN) (Batrinca & Treleaven, 2015); Recurrent Neural Network (RNN) (Goel et al., 2018), (Khan & Malviya, 2020); Single-layered Long-Short Term Memory (LSTM) (Moshkin et al., 2019), (Chintalapudi et al., 2021), (Mujahid et al., 2021), (Velampalli et al., 2022); and Neural Network (NN) (Mostafa et al., 2021), (Khasanova & Pasechnik, 2021), (Velampalli et al., 2022).

Table 1 Summary of the conclusions and observations of the reviewed systematization studies

Cite	Reviewed articles	Cited Methods	Conclusions/Contributions
(Drus & Khalid, 2019)	77 articles was screening and 24 selected for review. Period study: 2014 - 2019.	SVM, NB, TF-IDF (Das & Chakraborty, 2018), SentiWordnet (Agarwal et al., 2015)	– The obtained results demonstrated the usefulness of either Lexicon based method, Machine learning method or a mix of both methods when implementing sentiment analysis.
			– Most of the study adapted Sentiwordnet and TF-IDF method when conducting sentiment analysis.
(Sharma et al., 2020)	36 cited and 13 are summarized. Period study: 2013 - 2018.	Term presence versus term frequency, N-gram features, parts of speech tagging (Mejova & Srinivasan, 2011). Bayesian networks (Al-Smadi et al., 2019), SVM (Zainuddin & Selamat, 2014), ANN, Decision tree (DT) (Kotenko et al., 2015), Rule based classifier (Xia et al, 2016).	– High accuracy of classification depends upon the quality of selected features and classification algorithm used.
			– SVM and NB are used by the researchers as a reference model for comparing their proposed work.
			– Lexicon-based approach is used by the researchers to solve sentiment analysis problems as it is scalable and computationally efficient.
(Babu & Kanaga, 2022)	101 collected, 13 of applications in social networks are summarized. Period study: 2016 - 2020.	SVM, NB, RF, KNN, Convolutional Neural Network (CNN), Bidirectional LSTM (Bi-LSTM)	– Emoticons examples presented in a table form.
			– Summary of application articles that evaluate ML and Deep learning techniques on social networks data, where the accuracy obtained and the dataset used are highlighted.
			– RF and SVM achieved the best performance on Facebook and Twitter dataset with 84.6% and 93.1%, respectively.

Source: Own Elaboration

Table 2 Results of reviewed articles that apply and evaluate ML techniques for sentiment analysis

Cite	Methodology	Models/Algorithms	Dataset source	Dataset size (registers)	Best performance	Accuracy	Measure performance parameters
(Khasanova & Pasechnik, 2021)	Training a neural network to classify the text	NN and Bag of Words (BoW)	VKontakte	400000	NN	80 %	confusion matrix
(Chintalapudi et al., 2021)	A fine-tuned 12-layer Bidirectional Encoder Representations from Transformers (BERT) model	Logistic regression (LR), SVM, Single-layered LSTM, BERT and AdamW optimizer.	Twitter	3090	BERT	89 %	accuracy
(AlBadani et al., 2022)	ULMFiT combined with SVM and applied it on different sentiment analysis datasets	SVM, LSTM, ULMFit–SVM model	Twitter	28595	ULMFit–SVM model	99.78%	accuracy, training time, testing time
			Yelp	14485			
			IMDb	65000			
(Başarslan & Kayaalp, 2020)	Comparison of the SVM, ANN and NB algorithms with the TF-IDF and Prediction-based text representation (W2V) methods	NB, SVM, ANN, TF-IDF, W2V	IMDB	1000	ANN with W2V method	96 % and 86 % on IMDB and Twitter datasets respectively	accuracy, precision, sensitivity and F - score
			Twitter	4500			
(Alatabi & Abbas, 2020)	A sentiment analysis system built on Bayesian Rough Decision Tree (BRDT) algorithm	BRDT, DT	Facebook	4000	BRDT	99.62 % on Facebook dataset, 96.15 % on Movie reviews dataset	accuracy, precision, recall, F1 score
			Movie reviews	2000			
(Mostafa et al., 2021)	Comparison between different selected machine learning algorithms for classification	NB, SVM, KNN Classifier, LR, NN.	Twitter	448013	NN	81.33% (subject to dataset size)	accuracy, costs per iterations, memory usage, training time
(Carvalho & Harris, 2020)	The accuracy off-the-shelf technologies and the bag-of-words approach is studied	Sentiment analysis services provided by four major cloud platforms (IBM Cloud, Amazon Web	Twitter	14640	IBM NLU (Twitter dataset), Amazon Comprehend	85.4 % (IBM NLU), 76.3 % (Amazon Comprehend)	accuracy
			Facebook	3240			

Cite	Methodology	Models/Algorithms	Dataset source	Dataset size (registers)	Best performance	Accuracy	Measure performance parameters
		Services, Microsoft Azure, and Google Cloud)			(Facebook dataset)		
(Khan & Malviya, 2020)	A new approach and classification method	Rough Set Theory, PSO, NB, Hadoop based deep RNN (Başarslan & Kayaalp, 2023) method	Twitter	no se determina en el documento	Hadoop based deep RNN method	93.02 %	accuracy
(Moshkin et al., 2019)	Assessing of the sentiment analysis of social network texts through an original ontological method	Ontological method based on SentiWordNet, NB classifier, LR, SVM, LSTM network	VKontakte	420	NB	78 %	accuracy
(Karthika et al., 2019)	RF and simulated by using SPYDER	RF, SVM	Twitter	448013	RF	97 %	accuracy
(Goel et al., 2018)	Methodology for automatic multilingual processing	NB, RNN	Twitter	no se determina en el documento	RNN	96.15 %	accuracy, confusion matrix
(Velampalli et al., 2022)	Universal Sentence Embedding and S-BERT to embed sentences to sentence vectors.	NN, LSTM NN, S-BERT	Twitter	2253	S-BERT with Standart NN, Universal Sentence Encoder with LSTM NN	98 %	precision, recall, F - score and accuracy
(Cyril et al., 2021)	Automated learning with CA-SVM based sentiment analysis model	TGS-K means clustering, Semantic sentiment score (SSS), Gazetteer and symbolic sentiment support (GSSS), Topical sentiment score (TSS), CA-SVM based model	Twitter	1,000,000,000	CA-SVM based model	92.48 %	recall, accuracy
(Mujahid et al., 2021)	Analysis of the sentiments of people about e-learning.	TF-IDF and BoW, TextBlob, VADER, and SentiWordNet, CNN (Başarslan & Kayaalp, 2023), LSTM, CNN-LSTM, Bi-LSTM, Synthetic Minority Oversampling Technique (SMOTE) and Topic modeling with Latent Semantic Analysis (LSA).	Twitter	17155	DT, RF, SVM with TF-IDF, with SMOTE	95 %	accuracy, precision, recall, F1 score, confusion matrix and K-Fold Cross-Validation
(Rustam et al., 2021)	Feature extraction technique based on BoW and TF-IDF and a new methodology	RF, XGBoost classifier, SVM, Extra trees classifier (ETC), and DT, TF-IDF, BoW	Twitter	7528	ETC with concatenate BoW and TF-IDF	93 %	precision, recall, F - score, accuracy, confusion matrix
(Salmony & Faridi, 2021)	Negation scope identification methods to find negated tokens and investigate how these tokens can raise SA classifiers' accuracy. Presents a methodology workflow process that they used to conduct Twitter Sentiment Analysis.	NB, SVM, LR with BoW and TF-IDF vector representation technique	Twitter	1,600,000	LR with TF-IDF embedding negated token	81 %	accuracy
(Alharbi & El-kenawy, 2021)	A hybrid approach, named GWOPS, that combines Grey Wolf Optimizer (GWO) and Particle Swarm Optimization (PSO) algorithms, for training NN classifiers.	GWO, PSO and Genetic algorithm (GA) with NN.	Twitter	2000	GWOPS with NN	93,85 %	standard error of mean, standard deviation mean
(Hassan et al., 2021)	A compressive annotation guideline for manual coding of tweets.	SVM, LR, NB	Twitter	6388	SVM with uni-gram	85 %	accuracy, precision, recall, F1 score and ROC curve

Source: Own Elaboration

The compared supervised learning algorithms are SVM (Karthika et al., 2019), (Moshkin et al., 2019), (Başarslan & Kayaalp, 2020), (Chintalapudi et al., 2021), (Mostafa et al., 2021), (AlBadani et al., 2022), (Rahman et al., 2022); Clustering-based Adaptive SVM (CA-SVM) (Cyril et al., 2021); LR (Moshkin et al., 2019), (Chintalapudi et al., 2021); (Mostafa et al., 2021); BRDT (Alatabi & Abbas, 2020); DT (Alatabi & Abbas, 2020); NB (Başarslan & Kayaalp, 2020), (Goel et al., 2018), (Khan & Malviya, 2020), (Moshkin et al., 2019), (Mostafa et al., 2021), (Rahman et al., 2022); KNN (Mostafa et al., 2021) and RF (Karthika et al., 2019); combined in some cases with the natural language processing techniques like BoW (Carvalho & Harris, 2020), (AlBadani et al., 2022), (Rahman et al., 2022); TF-IDF (Rustam et al., 2021), (Salmony & Faridi, 2021), (Rahman et al., 2022); Word2Vec (Başarslan & Kayaalp, 2020), (Rahman et al., 2022), (Başarslan & Kayaalp, 2023); and PSO (Başarslan & Kayaalp, 2020).

Inside the approaches to sentiment analysis, the authors characterized and applied the Semi-Supervised Ultra-Lightweight Multi-lingual Fine-Tuning (SIS-ULMFiT) (AlBadani et al., 2022), Rough set theory (Khan & Malviya, 2020), Ontological method based on SentiWordNet (Moshkin et al., 2019). In general, in all approaches to the subject reviewed, it is recognized that operations like text preprocessing have a big impact on the accuracy of the system as this process facilitates dealing with words and the deletion of unwanted words. Also, feature selection is a significant process because it reduces the number of features by selecting the most useful set of features, decreasing the number of features means less training time.

The importance of data preprocessing operations is evidenced in the publications of Carvalho and Harris (2020) y Chintalapudi et al. (2021). The comparing of the off-the-shelf technologies around the sentiment analysis services provided by four major cloud platforms, like IBM cloud, Amazon web services, Microsoft Azure, and Google cloud, against the BoW approach; makes evident that the pre-trained models available on the aforementioned platforms are more accurate than the BoW approach. The difference between IBM NLU, the most accurate technology, and the BoW approach is more than 30 percentage points (Carvalho & Harris, 2020).

However, there is no single technology that is consistently more accurate than the others across different sentiments. For example, Amazon comprehend is highly accurate when classifying neutral and positive posts, but drastically less accurate when handling negative posts. The main features and pricing scheme of sentiment analysis services provided by major cloud platforms is a good contribution to the review of Carvalho and Harris (2020).

Likewise, the effect of types of text representation on the performance of sentiment analysis is investigated by Khan and Malviya (2020). They demonstrated the performance of SVM, ANN and NB algorithms with the traditional TF-IDF and W2V methods in two different datasets. The results of the experiments with TF-IDF on Twitter and IMDB datasets, showed that the ANN algorithm had the best performance with an accuracy of 89% and 86% respectively, while the results applying W2V on the same datasets, also showed the best performance in the ANN algorithm with an accuracy of 87% and 90% respectively. The NB gave the worst performance among others in both datasets and the applied experiments.

Assessing the sentiment of social network texts within a software System for Opinion Mining is developed through the research of Moshkin et al. (2019); the research proposes an original ontological method that takes into account the features of the text data presentation in social networks and develops an architectural scheme of the software. The best result was with the use of the NB classifier with a 78% of accuracy. The developed method based on the dictionary obtained from SentiWordNet showed an efficiency of 77%.

On the other hand, the ULMFiT-SVM model, proposed by AlBadani et al. (2022), introduces an effective deep learning architecture that combines the universal language model fine-tuning with a SVM. The extensive results on three real-world datasets (Twitter, Yelp, IMDb) demonstrate that the model increases detection efficiency and accuracy, being between 95.78% and 99.78%. Therefore, in the study, the sentiment analysis was restricted to document level and they did not consider the sentiment at the aspect level.

However, in (Cyril et al., 2021) the authors propose an Automated learning CA-SVM based sentiment analysis model, which has produced efficient results than other feature selection and classifiers such as ANN and k-means; their proposal achieved an accuracy of 92.48%. In spite of, when contrasting the previous results associated with the proposed models, better performance results are obtained compared to models that have already been pre-trained.

In the case of the novel deep-learning model called BERT, the accuracy of 89% is achieved. Despite this result, it is still superior to other models like LR, SVM, and LSTM, in the context of its application in the investigation of Chintalapudi et al. (2021). When contrasting the previous results associated with the proposed models, the performance is higher compared to models that have already been pre-trained.

Another interesting approach is the one developed by Khan and Malviya (2020), who proposes to allocate a review of real-valued input twitter data using deep RNN with Hadoop framework to distribute data for feature extraction process. The comparative performance between Hadoop based deep RNN method and the Rough set theory, PSO and NB methods showed that the proposed method is more accurate with the value of 93.02%.

In another hand, a singular application of sentiment analysis was reviewed about measuring the performance of Tweets that contain emojis. The investigation of Velampalli et al. (2022), aims to generate embeddings using Universal Sentence Encoder and SBERT sentence embedding models, to improve the classification accuracy using standard fully connected NN, and LSTM-NN models. Training the models using a distributed training approach instead of a traditional single-threaded model is implemented for better scalability. The text classification accuracy was almost the same for the NN and LSTM-NN models, around 98%. But, on the contrary, when the validation set was built using emojis that were not present in the training set, the accuracy drastically reduced to 70% (Li et al., 2023).

In the same way, the general comparison between algorithms applied on data from the social network for SA, the performance of the NN classifier algorithm, obtained in the investigation of Mostafa et al. (2021), stood out, were achieved the highest accuracy of 81.33%, followed by SVM, with 79%. The rest of techniques achieve an accuracy of less than 77%. However, the authors consider that the results are limited by the size of the selected dataset, so the work can be done on an even bigger data set to obtain better accuracy.

A better performance was obtained by using RF, with accuracy of 97%, among other algorithms like SVM, with the accuracy of 92%. The application of the RF algorithm also includes unbalanced data into the process and limits overfitting without increasing the error rate (Alatabi & Abbas, 2020). Meanwhile, in Karthika et al. (2019), went further into the construction of a special system and gave it the ability to classify each input as positive sentiment or negative sentiment employing a recently developed BRDT. The performance of the BRDT in two datasets (Facebook and movie reviews) was about 99.625% of accuracy.

Several architectures and methodologies are proposed by the authors cited in Table 2, being consistent with the approaches found in the literature for sentiment analysis applying ML techniques with data obtained from social networks. The use of Twitter is coincidental as the most used social network for obtaining data for sentiment analysis, due to the easy access, being Facebook, less consulted without intuitable or stated reason. But, most of the research work mainly focuses on tweets in only one language. A solution, in the form of a methodology, to multilingual sentiment analysis problems by implementing algorithms, was proposed by Goel et al. (2018) and it was complemented by comparing precision factors to find the best solution for this type of analysis.

The Google Translator API is used to translate the text into a common language, such as English, and then a sentiment analysis model is applied to the translated text to obtain a sentiment score. It has been cleared from the observations that, RNN classifier is significantly ahead of the other classifier in the task of predicting the feelings with almost an accuracy of 96%.

Keeping in view the adequacy and efficacy of machine learning models, the research of Mujahid et al. (2021), adopts TextBlob, Valence Aware Dictionary for Sentiment Reasoning (VADER) and SentiWordNet to analyze the polarity and subjectivity score of tweets text. Two feature extraction techniques, TF-IDF and BoW have been used to effectively build and evaluate the models. All the models have been evaluated in terms of various important performance metrics such as accuracy, precision, recall, and F1 score.

Performance comparison is carried out for results of TextBlob, VADER, and SentiWordNet, as well as classification results of machine learning models and deep learning models such as CNN, LSTM, CNN-LSTM, and Bi-LSTM. Results indicate that using the data balancing with SMOTE enhances the classification accuracy. DT, SVM, and RF perform very well and achieve an accuracy of 0.95 using Bow and SMOTE, while SVM achieves 0.95 accuracy using TF-IDF with SMOTE. VADER and SentiWordNet techniques are also used for performance comparison with TextBlob, and results indicate that TextBlob shows superior results for data annotation in comparison to VADER and SentiWordNet (Mujahid et al., 2021).

Finally, Rustam et al. point out another way to extract information and conform a dataset for sentiment analysis, by the use of identifiers provided by the IEEE dataport. The text of each tweet is extracted via IDs by an internal crawler using the Tweepy library (2021).

4. Challenges in the reviewed literature

Despite the large amount of research found regarding the application of SA with ML techniques in data obtained from social networks, challenges that the scientific community must face are still revealed. Optimizing hyperparameters is a critical problem in the development and design of an effective learning model for network sentiment detection. Problem that AlBadani et al. (2022) suggests solving, using a k-fold cross-validation strategy, which allows, according to these authors, to increase the overall performance of the single SVM to find the optimal RBF kernel parameters and to fine-tune the approach hyperparameters.

Another of the challenges faced during the sentiment analysis of Twitter data is the creation of noise while labeling the data and the major challenge lies in building technology that identifies and compiles the overall sentiment. To solve this, according to Khan and Malviya (2020), “a dedicated and integrated platform based on Twitter-based content is needed for extracting the obtainable information in public from huge text streams to synthesize and analyze the feedback of the customers” (p. 2).

Other aspects to take into account in the application of AS with ML techniques are those point out in the research of Cyril et al. (2021), where issues related to the effect of clustering in sentiment analysis are addressed, to classify a tweet under available class. of sentiment; the semantic sentiment score, to consider the semantic terms in tweets; the topical sentiment score, to measure the sentiment score according to the topical score; and the gazetteer and symbolic sentiment support (GSS), to represent the support of tweet towards any sentiment class according to the gazetteer and symbolic features present in the tweet.

6. Gratitude

ITESO, the Jesuit University of Guadalajara and CONACYT, México; for being part of the Doctorate in Engineering Sciences as a scholarship holder.

7. Financing

This work has been funded by CONACYT [Scholarship number 839008, 2022].

8. Conclusions

Various sentiment analysis methods with their performance parameters and methodologies for SA, have been explored in the present review and the reviewed literature shows that ML is a powerful tool to extract and process data from social networks and perform SA on a given topic. The approaches, methods and methodologies reviewed, provide a solution with a satisfactory degree of effectiveness in all cases, where high accuracy of classification depends upon the quality of selected features and classification algorithm used. The methodology to multilingual sentiment analysis provided by Goel et al. (2018), it's a reference for future experiments.

SVM and NB are recurrently used by the researchers as a reference model for comparing their proposed work. These two algorithms provide high accuracy with feature selection techniques. But, instead of relying on one method, studies have proven that combining both methods has better efficiency. Thus, in order to improve the outcome, it is recommended to combine both methods, lexicon and machine learning method, as it will complement each other, and the result is improved compared to using one approach only.

Based on the reviewed paper, twitter is the top social media platform used to collect information on user opinion. 85% of the reviewed papers, on this and the research reviewed in the cited studies, use twitter to collect information for sentiment analysis. In the same way, Facebook has the largest social media users in the world. But it is not very popular for sentiment analysis as the data is messy, it is not structured well, and people often use short forms and a lot of spelling errors. This makes the data harder to analyze.

The accuracy of the applied classification algorithms differs as a result of the size of the dataset, the topic being addressed, the complexity of the cultural and idiomatic lexicon, and the ML models with which the algorithm is combined. To the best of our knowledge and according our review, the ULMFiT model combined with SVM applied in the work of AlBadani et al. (2022), the BRDT developed by Alatabi and Abbas (2020) and the S-BERT with standard NN and the Universal sentence encoder with LSTM-NN explained by Cyril et al. (2021); are the algorithms that, according to the review carried out, have a better performance with an effectiveness between 98% and 99.8%.

For future recommendation, further investigation is needed to develop a universal model of sentiment analysis that can be applied to a different type of data, exploring other potential social networking sites to obtain users' opinion and expanding the context of sentiment analysis application.

9. References

Agarwal, B., Mittal, N., Bansal, P. & Garg, S. (2015). Sentiment Analysis Using Common-Sense and Context Information. Computational intelligence and neuroscience. 2015, ID 715730. https://doi.org/10.1155/2015/715730

Alatabi, H. A. & Abbas, A. R. (2020). Sentiment analysis in social media using machine learning techniques. Iraqi Journal of Science, 16(1), 193-201. https://doi.org/10.24996/ijs.2020.61.1.22

AlBadani, B. S., Shi, R. & Dong, J. (2022). A novel machine learning approach for sentiment analysis on twitter incorporating the universal language model fine-tuning and SVM. Applied System Innovation, 5(1), 13. https://doi.org/10.3390/asi5010013

Alharbi, M. S. & El-kenawy, E. M. (2021). Optimize machine learning programming algorithms for sentiment analysis in social media. International Journal of Computer Applications, 174 (25), 38-43. https://doi.org/10.5120/ijca2021921169

Ansari, M. T. J., & Khan, N. A. (2021). Worldwide COVID-19 Vaccines Sentiment Analysis Through Twitter Content. Electronic Journal of General Medicine, 18 (6), em329. https://doi.org/10.29333/ejgm/11316

Al-Smadi, M., Al-Ayyoub, M., Jararweh, Y., & Qawasmeh, O. (2019). Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features. Information Processing & Management, 56(2), 308-319. https://doi.org/10.1016/j.ipm.2018.01.006

Babu, N. V. & Kanaga, E. G. M. (2022). Sentiment analysis in social media data for depression detection using artificial intelligence: a review. SN Computer Science, 3 (74), 1-20. https://doi.org/10.1007/s42979-021-00958-1

Başarslan, M. S. & Kayaalp, F. (2020) Sentiment analysis with machine learning methods on social media. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 9(3), 5-15. https://doi.org/10.14201/ADCAIJ202093515

Başarslan, M. S. & Kayaalp, F. (2023). MBi-GRUMCONV: A novel Multi Bi-GRU and Multi CNN-Based deep learning model for social media sentiment analysis. Journal of Cloud Computing, 12 (5). https://doi.org/10.1186/s13677-022-00386-3.

Batrinca, B. & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. AI & Society, 30(1), 89-116. https://doi.org/10.1007/s00146-014-0549-4

Carvalho, A. & Harris, L. (2020). Off-the-Shelf Technologies for Sentiment Analysis of Social Media Data: Two Empirical Studies. 26th American Conference on Information Systems (AMCIS), 6. https://aisel.aisnet.org/amcis2020/social_computing/social_computing/6

Carvalho, J. & Plastino, A. (2021). On the evaluation and combination of state-of-the-art features in Twitter sentiment analysis. Artificial Intelligence Review, 54, 1887-1936. https://doi.org/10.1007/s10462-020-09895-6

Chandrasekaran, G., Nguyen, T. N. & Hemanth, J. D. (2021). Multimodal sentimental analysis for social media applications: A comprehensive review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11 (5), e1415. https://doi.org/10.1002/widm.1415

Che, W., Zhao, Y., Guo, H., Su, Z. & Liu, T. (2015). Sentence compression for aspect-based sentiment analysis. IEEE/ACM Transactions on audio, speech, and language processing, 23(12), 2111-2124. https://doi.org/10.1109/TASLP.2015.2443982

Chintalapudi, N., Battineni, G. & Amenta, F. (2021). Sentimental analysis of COVID-19 tweets using deep learning models. Infectious Disease Reports, 13(2), 329-339. https://doi.org/10.3390/idr13020032
Cyril C. P. D., Beulah, J. R., Subramani N., Mohan, P., Harshavardhan, A. & Sivabalaselvamani, D. (2021) An automated learning model for sentiment analysis and data classification of Twitter data using balanced CA-SVM. Concurrent Engineering, 29 (4), 386-395. https://doi.org/10.1177/1063293X211031485

Drus, Z. & Khalid, H. (2019). Sentiment analysis in social media and its application: Systematic literature review. Procedia Computer Science, 161, 707-714. https://doi.org/10.1016/j.procs.2019.11.174

Fersini, E. (2017). Sentiment analysis in social networks: A machine learning perspective. Elsevier, 91-111. https://doi.org/10.1016/B978-0-12-804412-4.00006-1

Goel, V., Gupta, A. K. & Kumar, N. (2018). Sentiment analysis of multilingual twitter data using natural language processing. 8th International Conference on Communication Systems and Network Technologies (CSNT), IEEE, 208-212. https://doi.org/10.1109/CSNT.2018.8820254

Hassan, S. U., Saleem, A. & Nawaz, R. (2021). Sentiment analysis of tweets through Altmetrics: A machine learning approach. Journal of Information Science, 47(6), 712-726. https://doi.org/10.1177/0165551520930917

Karthika, P., Murugeswari, R. & Manoranjithem, R. (2019). Sentiment analysis of social media network using random forest algorithm. 2019 IEEE international conference on intelligent techniques in control, optimization and signal processing (INCOS), IEEE, 1-5. https://doi.org/10.1109/INCOS45849.2019.8951367

Khan, M. & Malviya, A. (2020). Big data approach for sentiment analysis of twitter data using Hadoop framework and deep learning. 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), IEEE, 1-5. https://doi.org/10.1109/ic-ETITE47903.2020.201

Khasanova, A. M. & Pasechnik, M. O. (2021). Social Media Analysis with Machine Learning. 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), IEEE, 32-35. https://doi.org/10.1109/ElConRus51938.2021.9396713

Kotenko, I., Chechulin, A., & Komashinsky, D. (2015). Evaluation of text classification techniques for inappropriate web content blocking. 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), IEEE, 412-417. https://doi.org/10.1109/IDAACS.2015.7340769

Kung, J. Y. (2023). Elicit (product review). The Journal of the Canadian Health Libraries Association, 44 (1), 15. https://doi.org/10.29173/jchla29657

Li, X., Zhang, J., Du, Y., Zhu, J., Fan, Y. & Chen, X. (2023) A Novel Deep Learning-based Sentiment Analysis Method Enhanced with Emojis in Microblog Social Networks. Enterprise Information Systems, 17(5). https://doi.org/10.1080/17517575.2022.2037160

Mejova, Y., & Srinivasan, P. (2011). Exploring Feature Definition and Selection for Sentiment Classifiers. Proceedings of the International AAAI Conference on Web and Social Media, 5 (1), 546-549. https://doi.org/10.1609/icwsm.v5i1.14163

Moshkin, V., Yarushkina, N. & Andreev, I. (2019). The sentiment analysis of unstructured social network data using the extended ontology SentiWordNet. 12th International Conference on Developments in eSystems Engineering (DeSE), IEEE, 576-580. https://doi.org/10.1109/DeSE.2019.00110

Mostafa, G., Ahmed, I. & Junayed, M. S. (2021). Investigation of different machine learning algorithms to determine human sentiment using Twitter data. International Journal of Information Technology and Computer Science (IJITCS), 13(2), 38-48. https://doi.org/10.5815/ijitcs.2021.02.04

Mujahid, M., Lee, E., Rustam, F., Washington, P. B., Ullah, S., Reshi, A. & Ashraf, I. (2021). Sentiment analysis and topic modeling on tweets about online education during COVID-19. Applied Sciences, 11(18), 8438. https://doi.org/10.3390/app11188438

Ought. (11 de abril de 2023). https://elicit.org/

Page, M. J. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews, BMJ, 372 (160). https://doi.org/10.1136/bmj.n160

Rahman, H., Tariq, J., Masood, A., Subahi, A., Khalaf, O. & Alotaibi, Y. (2022). Multi-Tier Sentiment Analysis of Social Media Text Using Supervised Machine Learning. Computers, Materials and Continua. 74 (3). 5527-5543. https://doi.org/10.32604/cmc.2023.033190

Ray, S. (2019). A quick review of machine learning algorithms. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), IEEE, 35-39. https://doi.org/10.1109/COMITCon.2019.8862451

Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A. & Choi, G. S. (2021). A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. Plos one, 16 (2), e0245909. https://doi.org/10.1371/journal.pone.0245909

Salmony, M. Y. & Faridi, A. R. (2021). An enhanced twitter sentiment analysis model using negation scope identification methods. 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, 864-869. https://ieeexplore.ieee.org/document/9441151

SciSpace. (12 de junio de 2023). https://typeset.io/

Sharma, D., Sabharwal, M., Goyal, V. & Vij, M. (2019). Sentiment analysis techniques for social media data: A review. Conference on Sustainable Technologies for Computational Intelligence, Springer, 75-90. https://doi.org/10.1007/978-981-15-0029-9_7

Sharma, D., Sabharwal, M., Goyal, V. & Vij, M. (2020). Sentiment analysis techniques for social media data: A review. First International Conference on Sustainable Technologies for Computational Intelligence, Singapore, Springer, 75-90. https://doi.org/10.1007/978-981-15-0029-9_7

Sharma, Y., Mangat, V. & Kaur, M. (2015). A practical approach to sentiment analysis of Hindi tweets. 1st International Conference on Next Generation Computing Technologies (NGCT), IEEE, 677-680. https://doi.org/10.1109/NGCT.2015.7375207

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

Shofiya, C. & Abidi, S. (2021). Sentiment analysis on COVID-19-related social distancing in Canada using Twitter data. International Journal of Environmental Research and Public Health, 18 (11), 5993. https://doi.org/10.3390/ijerph18115993

Sumit, S. H., Hossan, M. Z., Muntasir, T. & Sourov, T. (2018). Exploring word embedding for bangla sentiment analysis. In 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), IEEE, 1-5. https://doi.org/10.1109/ICBSLP.2018.8554443

Velampalli, S., Muniyappa, C. & Saxena, A. (2022). Performance Evaluation of Sentiment Analysis on Text and Emoji Data Using End-to-End, Transfer Learning, Distributed and Explainable AI Models. Journal of Advances in Information Technology, 13(2), 167-172. https://doi.org/10.12720/jait.13.2.167-172

Xia, R., Xu, F., Yu, J., Qi, Y., & Cambria, E. (2016). Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis. Information Processing & Management, 52(1), 36-45. http://dx.doi.org/10.1016/j.ipm.2015.04.003

Xue, Z., Li, Q. & Zeng, X. (2023). Social media user behavior analysis applied to the fashion and apparel industry in the big data era. Journal of Retailing and Consumer Services, 72(103299). https://doi.org/10.1016/j.jretconser.2023.103299.

Yadav, J. (2023). Sentiment Analysis on Social Media. Qeios. https://doi.org/10.32388/YF9X04

Zainuddin, N., & Selamat, A. (2014). Sentiment analysis using support vector machine. 2014 International Conference on Computer, Communications, and Control Technology (I4CT), IEEE, 333-337. https://doi.org/10.1109/I4CT.2014.6914200