International Journal of Information Management Data Insights 3 (2023) 100180

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

A sentiment analysis framework to classify instances of sarcastic sentiments within the aviation sector

Abdul-Manan Iddrisu[a], Solomon Mensah[a,*], Fredrick Boafo[b], Govindha R. Yeluripati[b], Patrick Kudjo[c]

[a] Department of Computer Science, University of Ghana, P.O. Box LG 163, Legon, Accra, Ghana
[b] Department of Computer Science, Lancaster University Ghana, Ghana
[c] School of Computing & Technology, Wisconsin International University College, Ghana

ARTICLE INFO

Keywords:
Sentiment analysis
Social media
Aviation sector
Machine learning
Natural language processing

ABSTRACT

Social media in our current dispensation has become an integral part of daily routines. As a result, it is abundant in user opinions. Amid a global pandemic, these online platforms have taken a center stage in the disbursement of relevant information such as travel, emergency and pandemic hotspots. For researchers, this situation has presented itself as a challenge and opportunity to leverage big data for analysis and making informed decisions. This study seeks to develop a framework comprising of three operators, namely *Assemble+Deft, Edify+Authenticate* and *Forecast* to classify opinion instances as *sarcastic* or *non-sarcastic*. The framework is tested with a Twitter dataset using key state-of-the-art techniques, namely Recurrent Neural Network (RNN) with Gated recurrent unit and Support Vector Machines (SVM). The dataset consists of opinions on effect of COVID-19 pandemic on air travel. The evaluation metrics used include precision, accuracy, recall and F1-score. The experimental analysis showed a significant increase from 9.28% under a standard sentiment review to 10.1% optimized sentiment analysis. The findings further show a significant improvement in the performance of optimized SVM yielding an improved prediction performance compared to RNN. The outcome of this study will support airlines to understand the frustration and complaints of customers and to make concrete decisions on how to improve their services. The framework will serve as a benchmark for future sentiment analysis in other sectors where customer views and comments are core to their services.

1. Introduction

Since the first recorded case of the coronavirus disease 2019 (COVID-19) that was traced to a seafood market in the Wuhan province in the People's Republic of China in 2019 (Velavan and Meyer, 2020), the World Health Organization officially reported and projected over 303 million cases with a reported confirmed death of 5.48 million people (Zhao et al., 2020) over a few months.

COVID-19 had a detrimental influence on the free movement of people, especially in the airline industry which plummeted to an all-time low, leaving many passengers stranded at numerous airports. This led to widespread disruptions and changes in travel plans. Thousands of stranded airline passengers expressed their dissatisfaction about their predicament through sarcastic comments that they posted on social media platforms such as Twitter. Social media platforms have become a means companies and industries are able to connect to their customers to assess their dissatisfaction of services rendered (Gkikas et al., 2022). Sarcasm can be important to the airline industry with respect to sentiment analysis because it is a common form of language used by customers to express dissatisfaction or frustration, and the sentiment of a sarcastic statement is often the opposite of the literal meaning of the words used. Alongside the evolution of social media networks, the sheer volume of social media text available for sentiment analysis has increased multifold, leading to a formidable corpus (Kandasamy et al., 2020). A study by Grover et al. (2022) was conducted to investigate the impact of social media at the individual level with respect to different contexts such as organization, marketplace and social environment. In this study, 132 articles were selected for review and a conclusive outcome indicated that social media platforms and channels continue to influence public opinion as well as generate lot of data.

The most popular source of emotional data or opinions is from online platforms such as social media websites, blogs or live streaming services (Mehta & Pandya, 2020). Consumer engagement in social media brand communities plays a key role in defining how researchers can identify main conceptualizations and address the central social behavior and mass communication theories to support these relationships

* Corresponding author.
E-mail address: smensah03@ug.edu.gh (S. Mensah).

https://doi.org/10.1016/j.jjimei.2023.100180
Received 17 August 2022; Received in revised form 22 April 2023; Accepted 23 April 2023
2667-0968/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

(Santos et al., 2022). There are different techniques that can be used for performing sentiment analysis, however, the two most popular techniques include the use of machine learning in predictive scheduling and resource allocation in large manufacturing systems (Morariu et al., 2020) and lexicon-based which is associated with the textual method (Mehta & Pandya, 2020).

The objective of this research is to propose a framework that enhances the detection and analysis of sarcastic and non-sarcastic sentiments within the aviation industry. The heterogeneous task of detecting sarcastic opinion polls meant that there is inadequate research done in this perspective. It is important to consider sarcastic and non-sarcastic sentences in sentiment analysis because sarcasm is a common form of language used to convey irony or humor, and the sentiment of a sarcastic statement is often the opposite of the literal meaning of the words used. If a sentiment analysis system is not able to correctly identify and interpret sarcastic language, it may produce inaccurate results. As a result, this study intends to use sarcastic and non-sarcastic tweets to develop a classification model that can help detect and perform sentiment analysis in a much better way.

The outcome of this research allows us to effectively analyze the impact of COVID-19 on the aviation industry using the concept of sarcasm. Including both sarcastic and non-sarcastic sentences in the training data of a sentiment analysis can help the system learn to recognize and correctly interpret sarcastic language, which can improve the overall accuracy of the system and help the airline more accurately understand the sentiment of its customers.

This framework will be used to help stakeholders in the aviation industry understand how customers use sarcasm to express frustration about a situation or service, so that they can better serve their customers. This framework can be applied not just to the impact of COVID-19 in the aviation industry, but also to any area of concern for stakeholders in any service industry. It will allow stakeholders to interpret the motivations of their customers more accurately by "reading between the lines" of sarcastic comments.

Prior to this, we reviewed existing works in the field of sentiment analysis and the various state-of-the-art techniques currently used for sentiment classification. This serves as a guide to measure the efficiency of the proposed approach. Given the above, the current study outlines the following contributions to the research community:

(i) The study proposes a novel framework comprising of three operators, namely *Assemble+Deft, Edify+Authenticate* and *Forecast* to classify opinion instances as sarcastic or non-sarcastic.
(ii) To improve the detection and analysis of sarcastic and non-sarcastic sentiments within the aviation industry using Recurrent Neural Network (RNN) with Gated Recurrent Unit (GRU) and Support Vector Machines (SVM).

The rest of the paper is organized as follows: Section 2 presents preliminary concepts relating to the present work and previous research effort related to the concept proposed in the study. The research methodology is presented in Section 3. In Section 4, we provide details of the result from the empirical analysis. Section 5 provides the discussion of the result in relation to the contribution to literature and practical implication. Section 6 provides the study's conclusion and suggest future research directions.

2. Literature review

In this section, we will review exiting works in the field of sentiment analysis and how it has evolved over the years to serve various purposes. More recent approaches to sentiment analysis involve the use of machine learning techniques, such as natural language processing and deep learning. These techniques have been used to develop systems that can identify and classify the sentiment of a piece of text with high accuracy, taking into account the context and nuances of human language.

2.1. The evolution of data

In recent times due to the availability of large online data, sentiment analysis has become a critical tool in helping companies and stakeholders in various industries to understand and analyze the opinions and sentiments of their clients. The range of applications of sentiment analysis spreads from education to agriculture, transportation to marketing and product sales, etc. The highlight of this study is assessing the impact of the coronavirus on the aviation industry.

2.2. History of related sentiment analysis research

In the study by Kusyanti and Zakia (2019), sentiment analysis was used in designing an application for evaluating data using an updated version of the k-nearest neighbor technique. This feat was achieved by adopting a series of steps beginning from preprocessing to the calculation of cosine similarity (degree of similarity) and the training of data. In conclusion, developing a smart k-nearest neighbor technique resulted in the best average accuracy of 88.76%. Nonetheless, the best accuracy of all scenario tests was around 90.67% and an optimal k value of 15. By developing an improved k-nearest Neighbor technique, it was easier to review mobile application documents with positive and negative sentiments. One of the key features of sentiment analysis is the ability to extract knowledge related to opinions and emotions from users. In a study by Zucco et al. (2018), the concept of sentiment analysis was applied to the field of medicine under which various models of sentiments were experimented with. In their study, external explainer models such as rule extraction methods, attribution or relevance methods, and intrinsic methods were tested. However, during the implementation and analysis of the text sentiments, Long Short-Term Memory (LSTM) networks were employed to train the text and also to generate a description conditioned on the features extracted by the CNN modules.

Since the proliferation of e-commerce within the last two decades, there has been an overwhelming increase in the number of people shopping online. This has resulted in huge data being generated from these online platforms such as user preferences, reviews, ratings and many more (Yakubu & Kwong, 2020). Manufacturers have also taken advantage of this publicly available data to improve their market share and attract new clients by extracting and evaluating product reviews (Yakubu & Kwong, 2021). Data available on e-commerce platform such as Amazon.com was collected and used to perform sentiments on two levels of categorization namely, customer satisfaction and rating.

From the onset of the COVID-19 pandemic, social media has showcased a wide spectrum of people's perspectives and feelings, as well as associated incidents. Alamoodi et al. (2021) presented a comprehensive paper on how sentiment analysis and its application can be used in fighting COVID-19 and other infectious diseases. This was accomplished by first extracting textual sentiment from various social media platforms, including Facebook, Twitter, and Reddit. Secondly, a data collection procedure was used to obtain the desired information based on preferences. The third step was the pre-processing of extracted data and finally analysis of processed data. The results were then used for the intended purpose. Fig. 1 illustrates the entire process steps for sentiment extraction and analysis as defined by Alamoodi et al. (2021). The steps are data collection, extraction, pre-processing, data analysis and results.

2.3. Evolution of human sentiment analysis (Psychology perspective)

From a psychological perspective, the evolution of sentiment analysis technology has involved the development of methods for automating the process of identifying and interpreting human emotions and attitudes from text data.

Early approaches to sentiment analysis often relied on dictionaries of positive and negative words to classify the sentiment of a piece of text. However, this approach had limited accuracy and was not able to capture the full range of human emotions and attitudes.

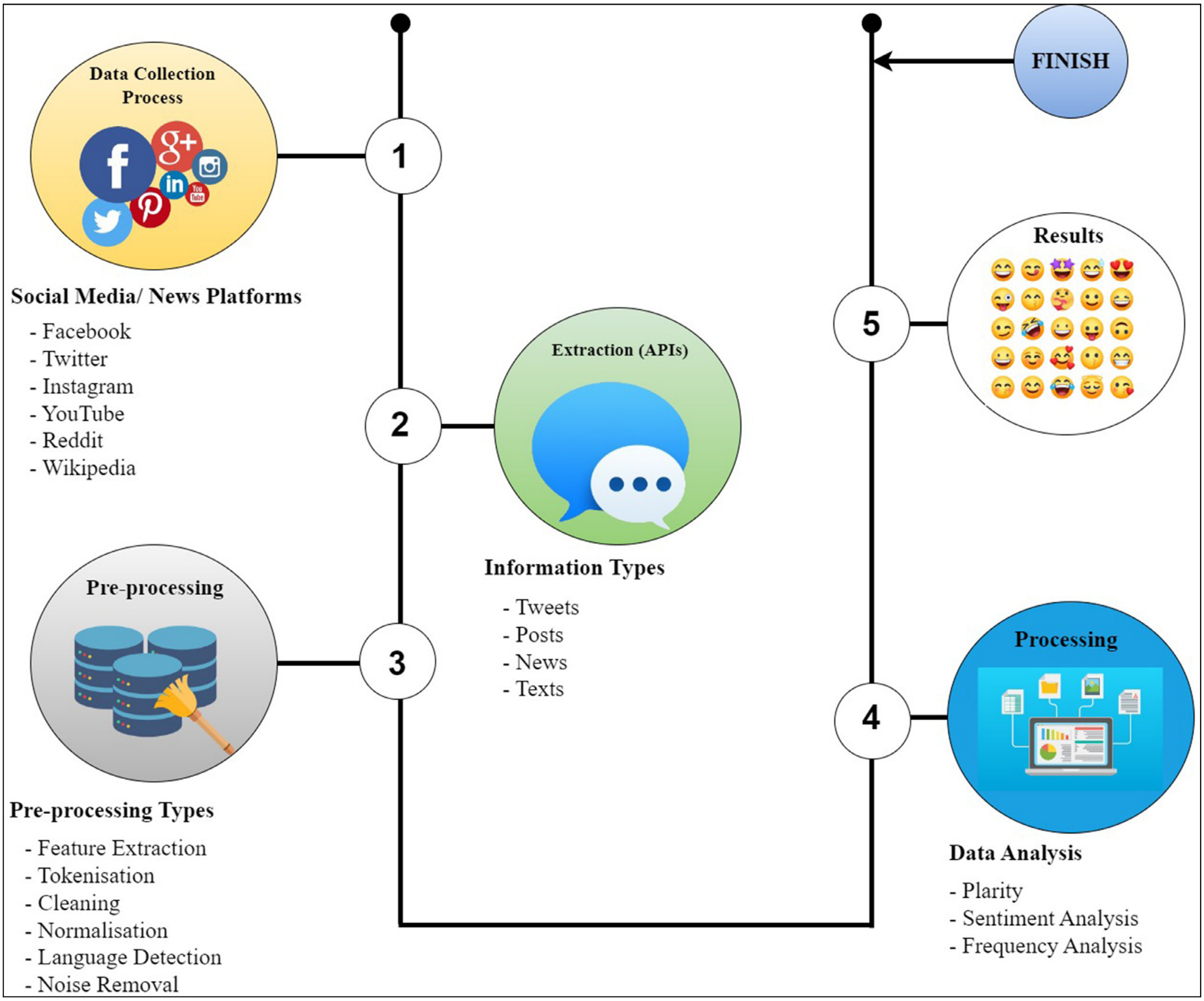

Fig. 1. Sentiment extraction and analysis steps (Alamoodi et al., 2021).

More recent approaches have involved the use of machine learning techniques, such as natural language processing and deep learning, which have enabled more accurate and sophisticated analysis of text data. These techniques have been used to develop systems that can identify and classify the sentiment of a piece of text with high accuracy, considering the context and nuances of human language.

There has also been a growing focus on developing methods for analyzing sentiment in real-time, such as using streaming data from social media platforms. This has enabled the development of systems that can track and analyze the sentiment of large groups of people in near real-time, providing insights into the emotions and attitudes of individuals and communities.

Overall, the evolution of sentiment analysis technology from a psychological perspective has involved the development of increasingly sophisticated methods for automating the identification and interpretation of human emotions and attitudes from text data, with a focus on improving accuracy and real-time analysis capabilities.

2.4. *Reviews on the implementation of classifications in sentiment analysis*

After collecting and reviewing datasets from 10 of the world's top social media platforms, (Hemmatian & Sohrabi, 2017) presented a framework of opinion mining to monitor, classify and distinguish between various aspect-based sentiment analyses based on the certified scientific methodology. Ahuja et al. (2019) proposed a methodology where six preprocessing techniques were applied to an SS-tweet dataset and extracted features using N-grams and text frequency-inverse document frequency (TF-IDF) techniques (Fig. 2). The number of times a term appears in a document divided by the total number of words in the document yields the term frequency (t).

Six (6) classification algorithms were suggested for by Ahuja et al. (2019) to be used for sentiment analysis and results evaluated using precision, recall, accuracy and F1-score as illustrated in Fig. 2. These algorithms work well with both categories as well as numerical data.

In a study by Chiarello et al. (2020), a new lexicon-based supervised learning method was proposed to filter consumer opinions from Twitter. According to a study by Avinash and Sivasankar (2019), the performance of feature extraction techniques, such as TF-IDF and document to vector (Doc2vec) was conducted using movie reviews from datasets of Cornell, UCI and Stanford. It was used to successfully classify texts into either positive or negative polarities by adopting various preprocessing methodologies such as removing stop words and tokenization. This technique improved the performance in terms of accuracy and processing time. Wang et al. (2021) proposed a sentence-to-sentence attention network (S2SAN) using a multiheaded self-attention model used to per-

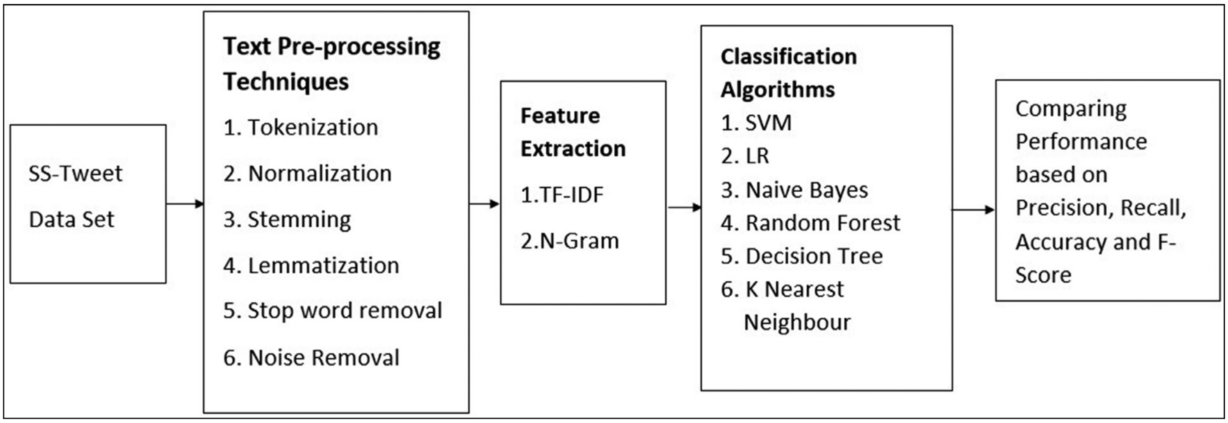

Fig. 2. Methodology for sentiment processing.

Table 1
Review of existing frameworks for sentiment analysis.

Name of Framework	Scope of Operation	Source of Dataset
Parsing-based Sarcasm Sentiment Recognition in Twitter Data. (Bharti et al., 2015)	In this paper, two approaches to detect sarcasm in the text of Twitter data were proposed. The first is a parsing-based lexicon generation algorithm (PBLGA) and the second was to detect sarcasm based on the occurrence of the interjection word.	Twitter
Sarcasm Detection on Twitter: A Behavioral Modeling Approach. (Rajadesingan et al., 2015)	This paper aims to address the difficult task of sarcasm detection on Twitter by leveraging behavioral traits intrinsic to users expressing sarcasm. Theories from behavioral and psychological studies were employed to construct a behavioral modeling framework tuned for detecting sarcasm.	Twitter
Sarcasm Detection on Facebook: A Supervised Learning Approach (Das & Clark, 2018a)	The use of user interaction pattern as a source of context information for detecting sarcasm. A supervised machine learning based approach focusing on both contents of posts (e.g., text, image) and users' interaction on those posts	Facebook
Sarcasm Detection in News Headlines using Voted Classification (see Fig. 3) (Bharti et al., 2022)	This paper deals particularly with sarcasm detection in News Headlines. The approach implemented is bag of words analysis using term frequency and n-grams frequency followed by voted classification. The study also outlines different approaches, namely supervised, unsupervised and semi-supervised techniques in the detection of sarcasm in a given text (Fig. 3).	News
Using LSTM for Context Based Approach of Sarcasm Detection in Twitter. (Khotijah et al., 2020)	The use of paragraph2vec to simplify the process of finding the contextual meaning that will provide the features to help classification in Long Short-Term Memory (LSTM).	Twitter
Sarcasm Detection on Flickr Using a CNN (Das & Clark, 2018b)	This paper presents a convolutional neural network-based model for detecting sarcasm based on images shared on a popular social photo sharing site, Flickr.	Flickr
Sarcasm Detection Using Graph Convolutional Networks with Bidirectional LSTM (He et al., 2020)	In this work, a new type of neural network model is proposed. Specifically, a graph convolutional neural (GCN) network is used to capture the features of global information in the satire context and jointly bidirectional LSTM (bi-LSTM) neural network to capture the sequence features of the comments respectively.	Reddit
Sarcasm Detection with Self-matching Networks and Low-rank Bilinear Pooling (Xiong et al., 2019)	Proposing a novel self-matching network to capture sentence "incongruity" information by exploring word-to-word interactions	Twitter

form sentiment analysis and it outperformed the existing state-of-the-art models. A generic framework was introduced by Kazmaier and van Vuuren (2020) to leverage opinion-bearing data to inform decision-making for sentiment analysis.(Table 1)

2.5. Comparative analysis of existing frameworks on sarcasm detection in sentiment analysis

Even though different frameworks have been developed and used to detect and measure sarcasm for the purpose of sentiment analysis, none of these frameworks have targeted a more practical human engaged services area driven on data.

Existing framework-based testing sentiment analysis does not provide robust and comprehensive comparative results of techniques to effectively detect and measure sentiment analysis. This study seeks to bridge the gap by identifying and using two state-of-the-art techniques, and develop a framework that will provide an improved result when used to test these techniques based on empirical data.

The proposed framework in this study specifically targets the aviation sector but can be replicated to research in similar environments where there is human engagement, and these engagements can be collected in the form of data. The proposed framework also employs the use of state-of-the-art sentiment techniques to test and compare the final output parameters such as precision, accuracy, recall and F1-score.

3. Methodology

For this study's sentiment classification, a Lexicon heuristic-based approach was used. Due to the distinct advantages and disadvantages of the two methodologies, namely Lexicon-based sentiment analysis and machine learning approaches, we merged the two methods for the purpose of this study. Both techniques have the advantage of being solely text-based and as such eliminates other complex analysis processes. However, with machine learning, various sub-techniques can be employed to undertake sentiment analysis. The preferred data extraction tool used for this study is Postman with Twitter API version 2. Because it is the simplest and clearest variable element in social media analysis, Saura et al. (2018) User-generated Content (UGC) is implemented and used. The growth of UGC, combined with the development of analytical technologies like big data, data mining and machine learning, has

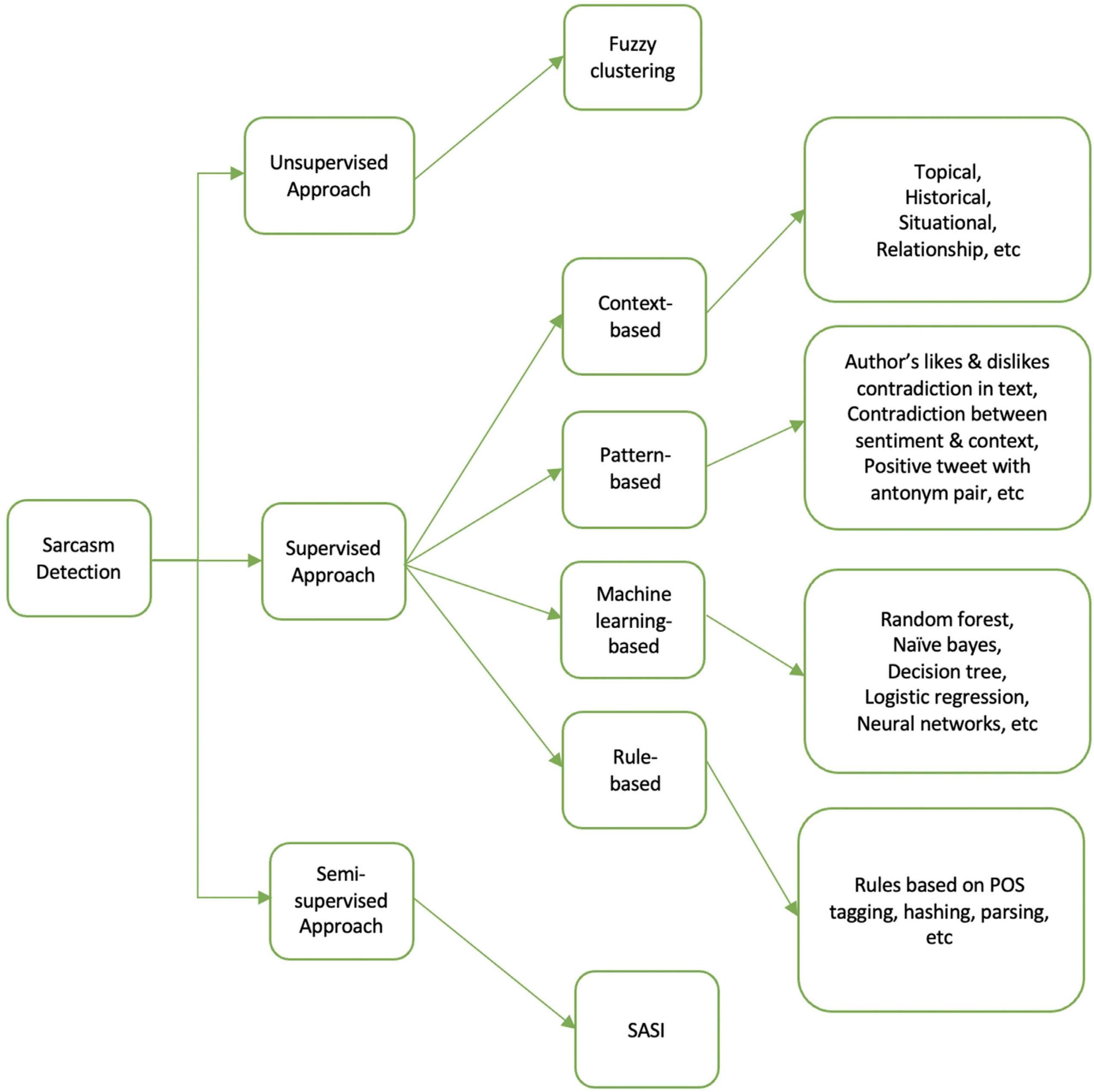

Fig. 3. Classification Methods for Sarcasm Detection.

resulted in a plethora of data optimization methodologies for UGC analysis. This has also been used in the digital tourism services (Kitsios et al., 2022) as generated content becomes valuable to other users. A study by Kumar et al. (2021) explored various text mining application in services and management. It revealed that such applications covered areas such as hospitality, information processing and management. The major purpose of this analysis is to identify key indicators (KIs) that can assist businesses in making better strategic decisions in the digital world. The KIs used in the sentiment lexicon library include *late, boarding gate, airport, staff, pilots, unprofessional, anxiety, chaos, meals, refunds, aircraft change, fare terms, aircraft seats, cabin cleanliness, hidden costs, customer service, failure, charges, poor booking, upgrades, nose mask, testing, cabin washrooms, onboard service, ticket flexibility* and *coronavirus*. Other factors such as modeling content readability, length, and hashtags number play a key role in determining how text characteristics impact user engagement through social media (Gkikas et al., 2022). Fig. 4 summarizes the KIs used for the text mining classification of a given sentiment as sarcastic (positive) or non-sarcastic (negative).

3.1. Description of dataset

Many existing social media platforms create a large amount of user-generated content, which has been used in many studies such as market analyses and online surveys. This study is primarily focused on evaluating text-based data; hence we chose a social media platform, namely Twitter which is more text-centric and the largest microblogging platform with over 260 million monthly active users and 500 million tweets per day. This microblogging platform is regarded as a trustworthy and credible channel to disburse information (El Rahman et al., 2019). US sentiment tweets are extracted from Kaggle[1] for this study.

[1] https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment.

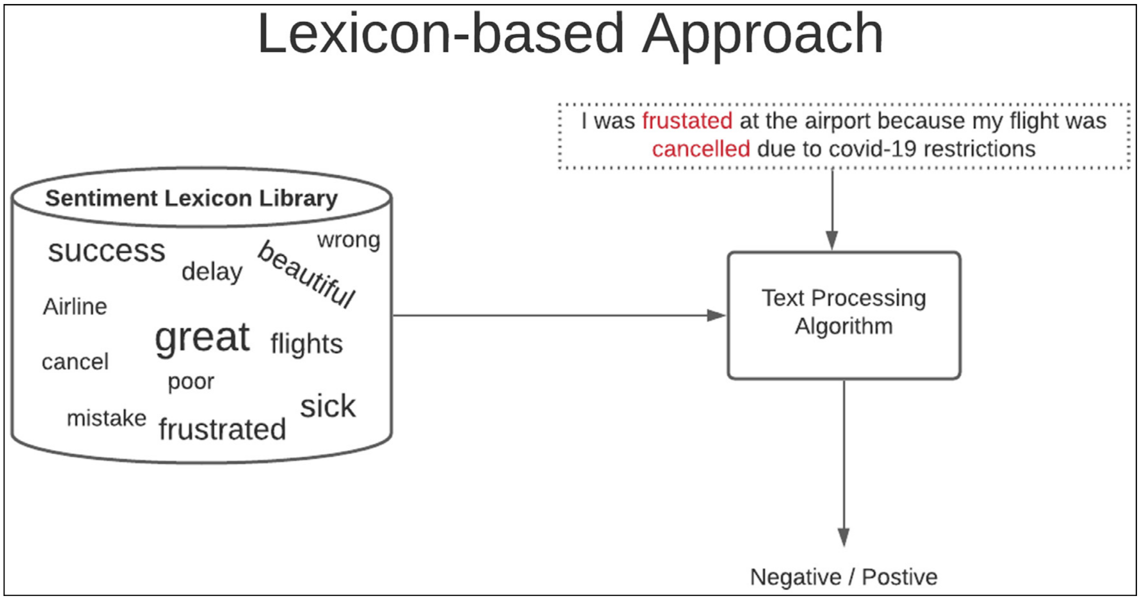

Fig. 4. Lexicon-based text processing approach.

These tweets were scrapped and analyzed from across the US airline industry and comprised six major airline carriers. The expression of negative sentiments turn to drive or encourage the spread of sentiments hence turns to have a higher impact on final outcome of any sentiment analysis (Sufi, 2022). This contains comments from passengers based on the service provided by airlines. As of the time of conducting this study, the latest update to the data was made during the COVID-19 pandemic to reflect the comments of passengers. However, this data has been reformatted to fit the purpose of this study. The data originally came from Appen which is a leading global machine intelligence company.

Natural language processing (NLP) which has been implemented for sentiment classification has aided in the processing of a large quantity of textual data (Kandasamy et al., 2020). In a study by Da Silva et al. (2014), classifier ensembles were used for tweet sentiment analysis. Similar modules were used by Chan and Chong (2017) for a study in determining sentiment analysis in financial texts.

The following metrics were used: *user id, text, tweet creation date, and public metrics like retweet count, reply count, like count* and *quote count*. The search string was dynamically updated depending on the amount of data obtained during the data extraction process. The data was extracted and saved in a JSON file. This was later converted to a CSV.

In Fig. 5, the pseudocode shows key indicators used in the extraction of the tweet. This includes the tweet user id, text contained in the tweet, date, and time the tweet was created. The public metrics show a breakdown of how other users on Twitter reacted to the main tweet. This has been broken down to show the number of retweets, reply count, like count and quote count.

The Algorithm Development Process:

(1) Identify the tweet using the id number
(2) Select a tweet that contains keywords from the sentiment lexicon library
(3) Indicate the date and time of the creation of the tweet
(4) Create a public metrics count
(5) Indicate the number of retweets
(6) Indicate the number of tweets reply count
(7) Indicate the number of tweets like count
(8) Indicate the number of tweet count
(9) Indicate the number of tweet quote

Fig. 6 provides a code fragment for extracting data from Twitter based on the above algorithm.

3.2. General data pre-processing

After extraction of the data, it was important to sanitize the raw data and prepare it for processing. The steps involved in this process can be categorized into four main steps. These are: *Data Cleaning, Data Reduction, Data Transformation* and finally *Data Integration.*

3.2.1. Procedure for tweet data processing

As a measure to ensure data integrity, several key steps were used and strictly complied with to attain this goal. These five steps are illustrated in Fig. 7. That is, separation and removal of retweets, removal of repetitive tweets, unrelated tweets, tweets with commercial accounts, and commercially related tweets.

3.3. Current state-of-the-art techniques in sentiment analysis

From the time Web 2.0 was introduced, the number of blogs, social media platforms, and forums has exploded allowing users to discuss and share their thoughts online. Many applications such as recommender systems, corporate survey analysis, and political campaign preparation all rely on this type of user data. In an ever-growing and competitive environment, businesses have now placed premium importance on the views and feedback of customers regarding a specific product or service. In studies by Cambria et al. (2017) and Hussein (2018), the conclusive remarks were that, testing the ideological and methodological frameworks behind sentiment analysis is a challenge with varying results in terms of accuracy and efficiency. According to Bhavitha et al. (2017), there are three techniques for addressing the problem of sentiment analysis at the moment, namely (1) hybrid technique, (2) lexicon-based (wordbook) approach and (3) machine learning.

In classifying sarcasm under sentiment analysis, emphasis is laid on additional contextual information such as the tone of the comment or the relationship between the speaker and the subject. Some natural language processing (NLP) models have also been developed that are specifically designed to detect sarcasm in text. Ri et al. (2021) adopted a hybrid methodology of deep learning to detect sarcasm. These models

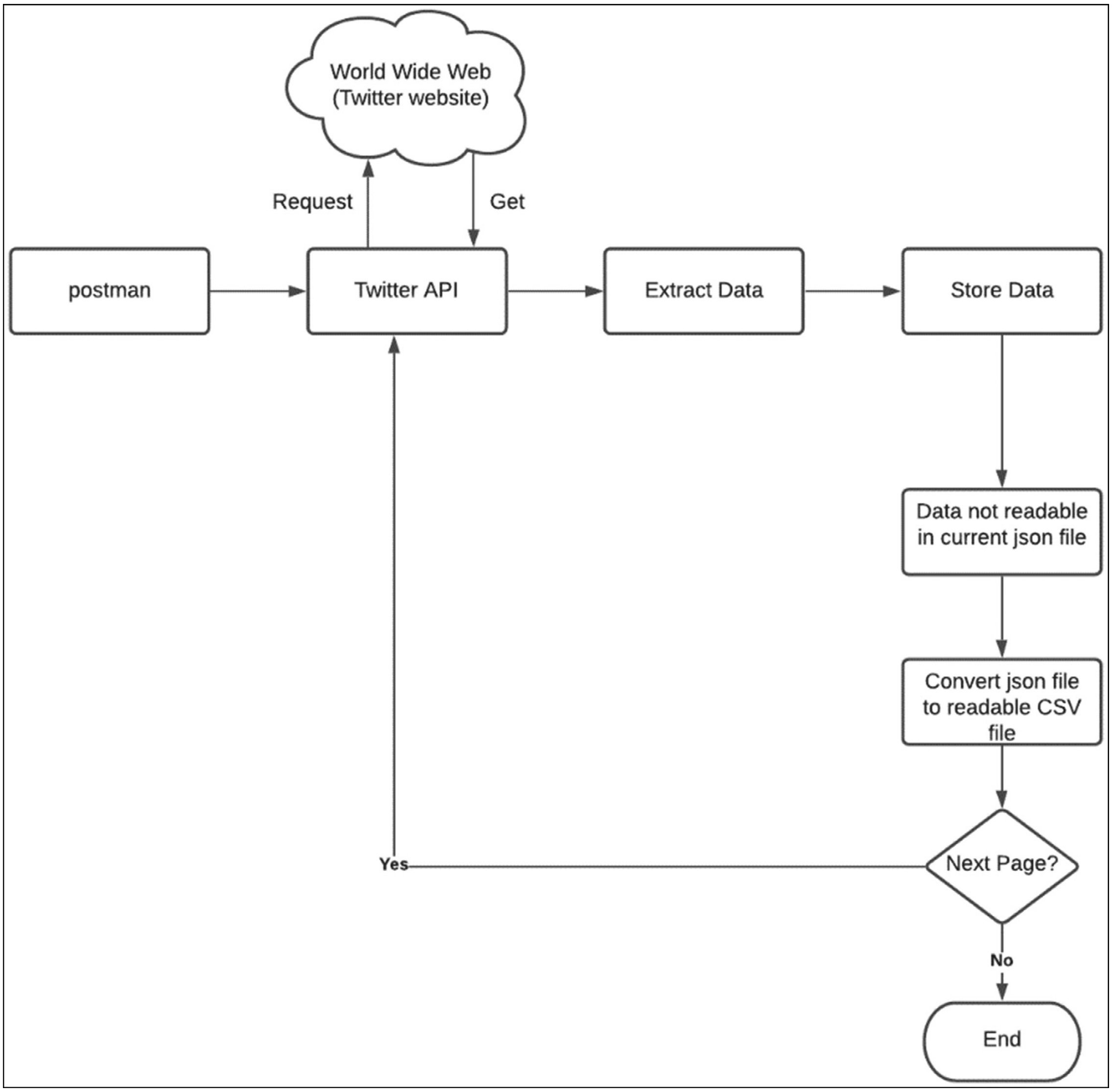

Fig. 5. Data Extraction Process using Postman.

often use machine learning techniques to analyze the words and structure of a sentence to determine whether it is likely to be sarcastic or not. In this study, the determination of sarcasm or sarcastic sentences, non-sarcastic and neutral sentences was done manually by an expert in the field of research. Once these sentences have been identified, they are used as part of a training dataset for a sentiment analysis model. The model can then learn to identify sarcastic and non-sarcastic sentences based on patterns and features that it extracts from the training data.

3.3.1. Machine learning (ML) vs other models

Within the last decade, researchers have developed and proposed different techniques to simplify and streamline the utilization of big data. These are primarily used in a wide range of applications such as determining the price of stock markets, real estate pricing and product review as used in studies by Jangid et al. (2018); Sohangir et al. (2018) and its application in the medical field as shown in the works of Satapathy et al. (2017). According to studies by Abid et al. (2019), Alharbi and de Doncker (2019) and Li et al. (2018), a limited number of research works laid emphasis on developing a technique that experimented with combining different deep learning algorithms to measure their performance. Based on value co-creation for open innovation, an evidence-based study of the data driven paradigm of social media was conducted using machine learning (Adikari et al., 2021). We use a combination of deep learning and conventional machine learning algorithms to examine the efficiency, precision, and accuracy of a proposed framework which was validated empirically using the Twitter dataset used in this study.

Fig. 8 presents a taxonomy of various methods and techniques that can be used for sentiment analysis. Techniques for sentiment analysis can be categorized into two main categories, namely Lexicon-based and Machine learning (Aloqaily et al., 2020; Britzolakis et al., 2020; Khaleghparast et al., 2023). A hybrid method can be obtained by combining methods from both Lexicon-based and Machine learning (Bhavitha et al., 2017). Machine learning algorithms can be classified into supervised, unsupervised, semi-supervised and reinforcement learning (Kar et al., 2022). Examples of the algorithms for sentiment analysis are illustrated in Fig. 8.

3.3.2. Deep learning (DL)

The hidden layers of a neural network are treated with a multilayer approach in deep learning (DL). That is, the deep learning architecture has multiple hidden layers each with at least one neuron. Advancements in the field of deep learning techniques and artificial neural networks present the most promising solutions to barriers faced with the processing of text base, image, and audio data.

```
"data": [
    {
        "id": "1345120767504220160",
        "text": "U.S. airlines have protocols intended to protect passengers from the coronavirus. \n\nBut how
            often people with COVID-19 board planes is impossible to know. \n\n@hugomartin reports: https://t.
            co/Q9BlqlVkVl",
        "created_at": "2021-01-01T21:32:36.000Z",
        "public_metrics": {
            "retweet_count": 1308,
            "reply_count": 366,
            "like_count": 1744,
            "quote_count": 563
        }
```


Fig. 6. Code fragment for extracting data from Twitter.

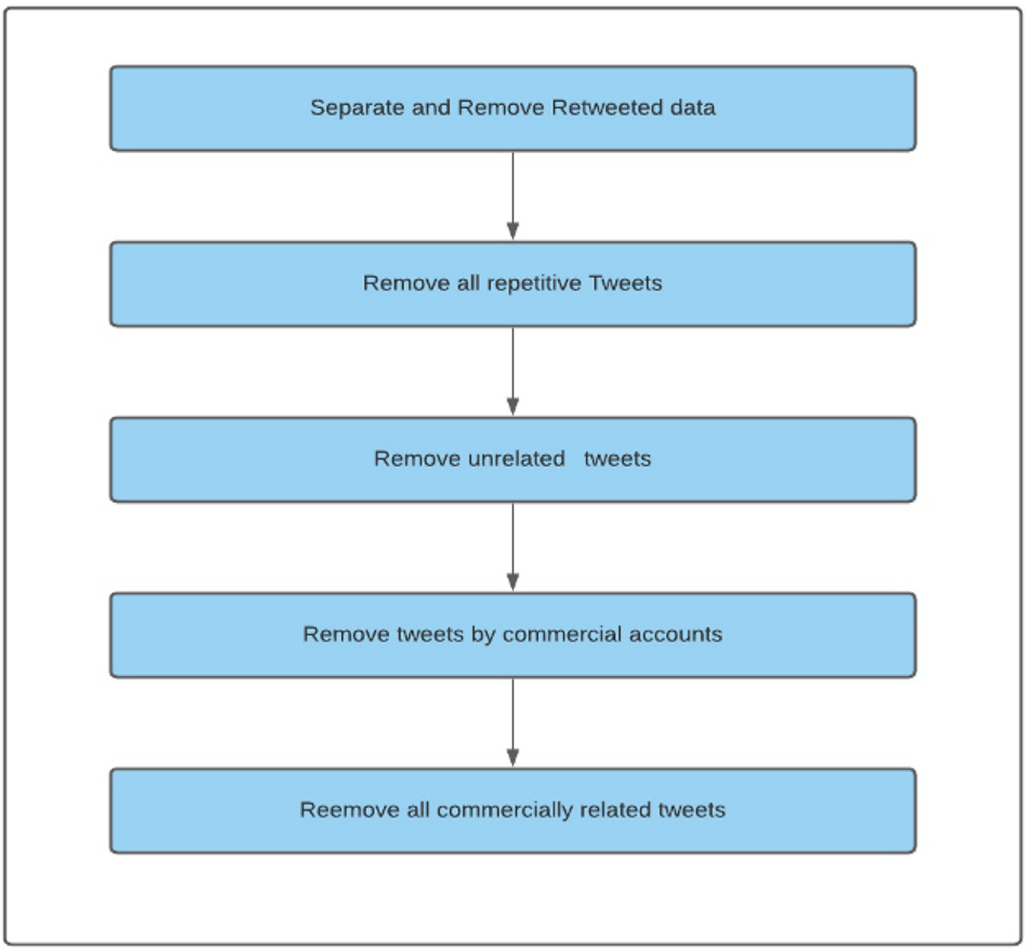

Fig. 7. Procedure for preprocessing of tweets.

When analyzing visual, audio, and natural language data, adopting a deep learning approach produces a relatively accurate result. Fig. 9 presents a deep learning architecture for analysing and classifying a given sentiment into the positive or negative class respectively (Bhavitha et al., 2017).

In a study by Abid et al. (2019), the performance of both Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) was tested, and it was determined that they both exhibited above-average overall accuracy when assessing them using a common strategy on a specific dataset inside a dedicated domain. An experiment conducted by Hassan and Mahmood (2017) also revealed that deep learning employing CNN and RNN models conquered barriers associated with a short text. The performance of Long Short-Term Memory (LSTM) was above satisfactory levels when tested on different tweets expressing several temperaments (Qian et al., 2018). A review of studies within the sentiment classification space indicates that most studies focus mainly on measuring features such as accuracy and F1-score at the detriment of other key factors such as processing cycle, precision and recall. Other limitations include using a relatively smaller dataset for analysis. This study seeks to provide alternative solutions by using a large dataset from Twitter and expanding the scope of classification by using different supervised and unsupervised techniques. The methodology and preceding

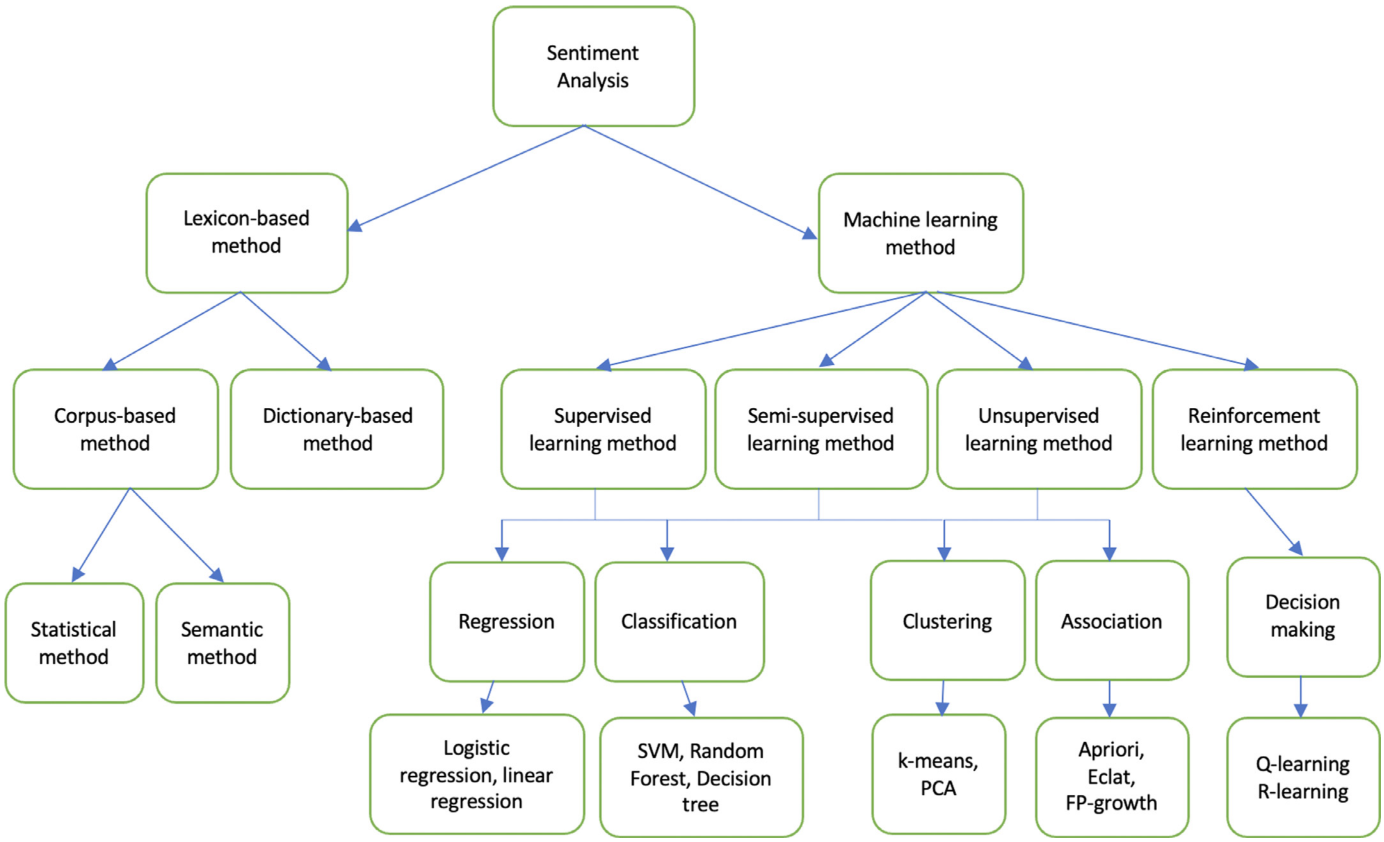

Fig. 8. Taxonomy of Sentiment Analysis Methods.

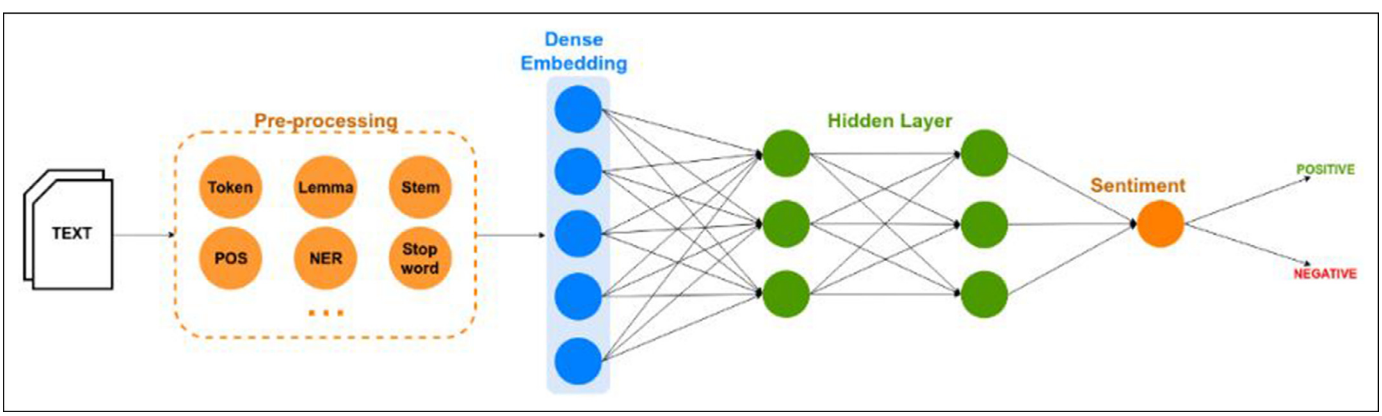

Fig. 9. Structure of a deep learning model.

experiment tested metrics such as confusion matrix, a measure of accuracy and precision.

We initialized the review of deep learning by conducting a comparative study of already existing techniques. We used a deep learning approach, namely an RNN with Gated Recurrent Unit (GRU) and a conventional machine learning approach, namely Support Vector Machine (SVM). The dataset used to evaluate the performance measures of these two approaches are based on the COVID-19 pandemic and were obtained from Twitter.

Recurrent neural network model. The structure of an RNN is indistinguishable from a feed-forward system, but it connects units in the same layer. This allows them to store input sequences of varied durations in internal memory. It comes with sequences of varying lengths by activating a recurrent hidden state whose performance relies on the activation of the previous time. Because the human brain is structured to operate like an RNN or a system of neurons with response attachments, engaging them is the most natural design we can achieve. The predictive nature of RNN has been exploited in the industry to address complex predictive tasks by leveraging the increased availability of data from processes (Brusaferri et al., 2020). The gradients either become too small to learn or become too large, leading the weights to surpass the maximum limit. The most effective solution to this problem is to incorporate a gating mechanism within the RNN. The basic role of RNN is to process sequential input using internal memory gained during guided cycles. A basic architecture of a traditional RNN is presented in Fig. 10 (Brusaferri et al., 2020).

For the purpose of sentiment classification, the many-to-one RNN classification type was used (Fig. 11). This model can collect raw text and embed it using tokenization. This is then trained to determine the

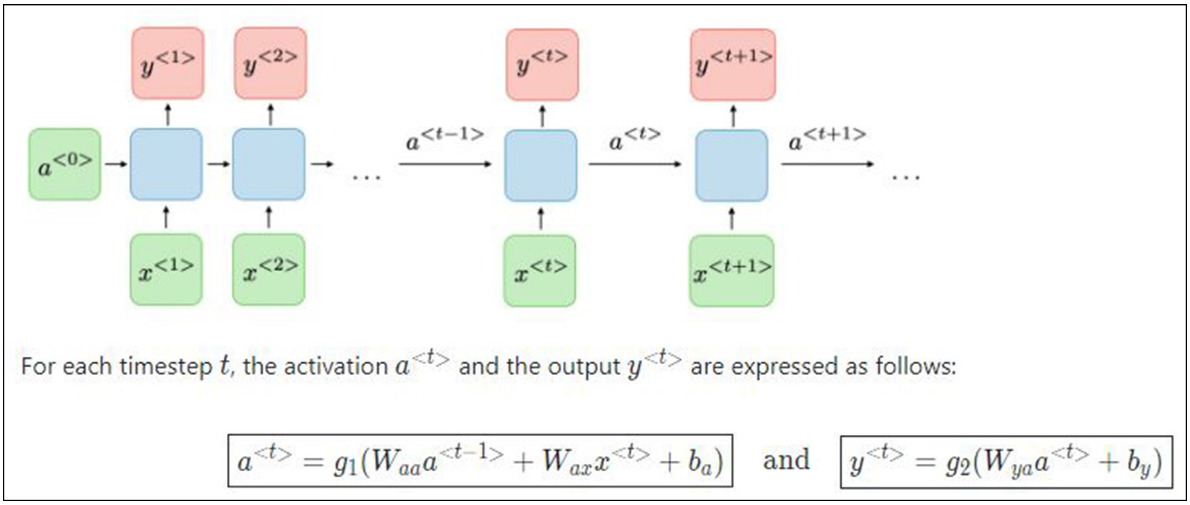

Fig. 10. Architecture of a traditional RNN.

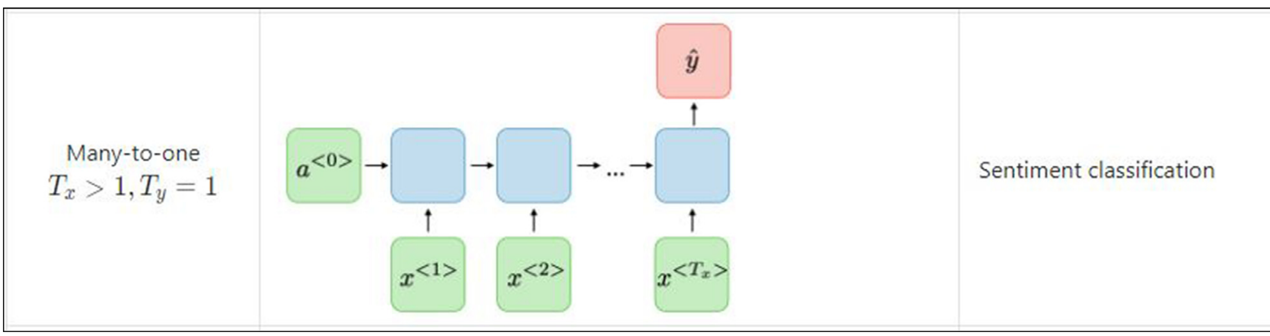

Fig. 11. Architecture of a many-to-one RNN.

type of sentiment being used. It should be noted that RNN has been recommended has one of the best Deep learning models for sentiment analysis when using data extracted from social media sites (Sufi, 2022).

Recurrent neural network (RNN) with gated recurrent unit (GRU) network architecture. In a bid to circumvent the vanishing gradient problem of a normal RNN, a gated recurrent unit uses the so-called update gate and reset gate. In essence, two vectors are used to select which data should be transmitted to the output. They are also capable of being trained to store and recall relevant data necessary for use in forecasts or other predictive analyses. This helps to save time and resources without needing to repeat training steps for data acquisition. In a study by Simeon (2017), this is described in greater depth. The structure of the RNN with GRU network architecture is presented in Fig. 12.

3.3.3. Support vector machine (SVM)

The SVM concept was first developed in 1995 by Cortes et al. (1995) as a machine learning technique for classification or categorization. For textual polarity detection, this is one of the most effective and extensively used supervised machine learning algorithms. SVMs are commonly referred to as universal learners. They have the unusual property of being able to learn independently of the feature space's dimensionality. The margin that divides the plane, rather than the number of features, is used to determine the complexity of a hypothesis (Joachims, 2019).

Text classification using the SVM learning algorithm. The output integer is an integer between 0 and 1 when the SVM input and output formats are established. While the input is a vector space, the output is a vector space (positive or negative). For a machine learning system to understand and process traditional text documents, it must first be converted to a format in which it can be used. During the process of conversion, each word will have a dimension allocated to it, and identical words will have the same dimension. High-dimension input space and document vector space are two of the advantages of employing SVM.

Evaluation of SVM. To identify which text classifier is better, performance measurements are utilized to evaluate different text classifiers. Some measure performance in a single binary category, while others use a combination of per-category measurements to provide an overall score. True positive (TP), false positive (FP), true negative (TN), and false negative (FN) obtained from a constructed confusion matrix reflect the number of true or false positives or negatives. As demonstrated below, the precision can be determined using the TP and FP rates:

$$Precision = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}) \tag{1}$$

TP stands for sentences that are correctly classified, while FP stands for sentences that are incorrectly classified.

Recall can be enumerated as:

$$Recall = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \tag{2}$$

FN is designated for non-classified sentences and TP represents correctly classified sentences

Formulae for F-measure are computed as:

$$F - measure = (\mathrm{Precision} * \mathrm{Recall} * 2)/(\mathrm{Precision} + \mathrm{Recall}) \tag{3}$$

3.3.4. Comparative analysis between ml and dl techniques

To establish a comparative analysis between ML and DL techniques, three successfully conducted research works were used to assess the efficacy of these two techniques. In a paper by Jain and Kaushal (2018) a

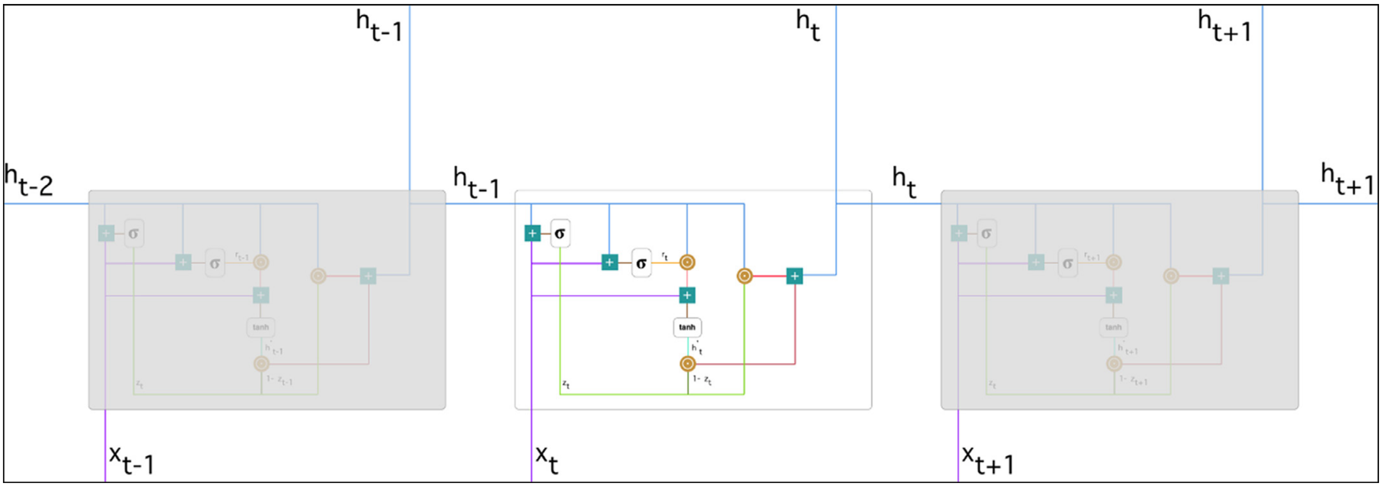

Fig. 12. Structure of RNN with GRU Network Architecture.

study was conducted to compare various ML, DL as well as hybrid techniques to measure their accuracy for sentiment analysis. It was concluded that in most cases DL techniques gave better results. In some rare cases, however, the difference in the accuracies of the two techniques is not substantial enough and, in such cases, the usage of ML was much easier while the DL method only increases the complexity. A similar comparative analysis of ML and DL was used to predict post-induction hypotension in patients after surgery (Lee et al., 2020). In their research two different types of algorithms from the DL category were adopted which included CNN and DNN whiles two traditional ML techniques namely Random Forest (RF) and Xgboost were also used. Among the predicted models tested, the RF, an ensemble tree, showed the best performance using the statistical features on vital records, and the CNN model showed the second-best performance using raw vital records. (Wang et al., 2021) also evaluated image classification algorithms based on traditional ML and DL using SVM and CNN respectively. It was concluded from their research that SVM gave an accuracy of 0.88 while CNN gave an accuracy of 0.98 when using a large sample dataset; when using a small sample COREL1000 dataset, the accuracy of SVM was 0.86 and the accuracy of CNN was 0.83.

3.4. Framework for detecting sarcastic sentiment

The proposed framework for detecting sarcastic sentiment comprises of three operators, namely *Assemble + Deft Assessment, Edify & Authenticate, and Forecast of New Prototypes*. Details of the operators are presented as follows:

3.4.1. Assemble + deft assessment (ASSEMBLEDEFT)

After classifying the dataset from the data bank, we utilize the X-means clustering method to create various segregation that satisfies the Bayesian Information Criterion. Different clusters are created using the X-means clustering method, and then stratification is achieved using these groupings. Each cluster is made up of a collection of data that is arranged chronologically but has no dates. The grouping is important because it allows us to discriminate between sarcastic and non-sarcastic attitudes. Our clustering results are then verified using expert judgment based on human intuition. That is, judgment was done by the authors of this paper. These two processes provided support in creating precise labels for the training and validation of the models. For rendition, we use variables zero (0) and one (1) to stand for sarcastic and non-sarcastic opinions.

3.4.2. Edify and authenticate

We may proceed with training and validation for an unbiased evaluation utilizing a supervised learning method once we have data in nice form sets of features. Following the compilation of training and validation, data loaders are required to batch the dataset and apply it to a sentiment network. In our scenario, the RNN layer, which uses a given hidden state size and the number of layers to turn the tokens into a certain embedding is used. A fully linked output layer maps the layer's output to the proper size, and a sigmoid activation layer performs the conversion. The well-formed and labeled dataset are used to create training, validation and test sets. As a result, the split fraction denotes the percentage of data to maintain in the training set. In contrast to other values such as 30%, 40%, or 50%, we utilized 80% as an experimental figure, which gave us accurate findings without any issues such as overfitting. To construct the validation and testing data, the remaining data was split in half, that is 10% for validation and 10% for testing.

3.4.3. Forecast of new prototypes

As part of the functionality of this framework model, it will be able to forecast any new outcome using its predictive feature and therefore give a final verdict on whether an expressed opinion is sarcastic or not. This operator makes use of the data execution technique embedded into it to predict instances of sarcastic and non-sarcastic sentiment depending on what it learns during the training process. To assess the prediction values, we estimate the median deficit and confusion matrix then use metrics such as precision, time loss and conditioning time, and then use them to categorize each case of the opinions into either sarcastic or non-sarcastic.

We present the flowchart and pseudocode for the proposed framework in Figs. 13 and 14, respectively.

4. Results

4.1. Data descriptive summary

The dataset is divided into two parts: an unlabeled dataset collected from Twitter and a labeled dataset from Kaggle. The labeled dataset was analyzed using a variety of feature extraction approaches such as the Principal Components Analysis (PCA) and Independent Component Analysis (ICA) which take as input data a mixture of independent components and it aims to correctly identify each of them (deleting all the unnecessary noise). PCA works by taking an original input and trying to find a combination of the input features which can best summarize the original data distribution. These methods provided support in making the raw texts more intelligible. After the analysis, three main data description classes were extracted, namely sarcastic, non-sarcastic and neutral (Table 2 and Fig. 15). However, for this study, sarcastic and non-sarcastic sentiments were used.

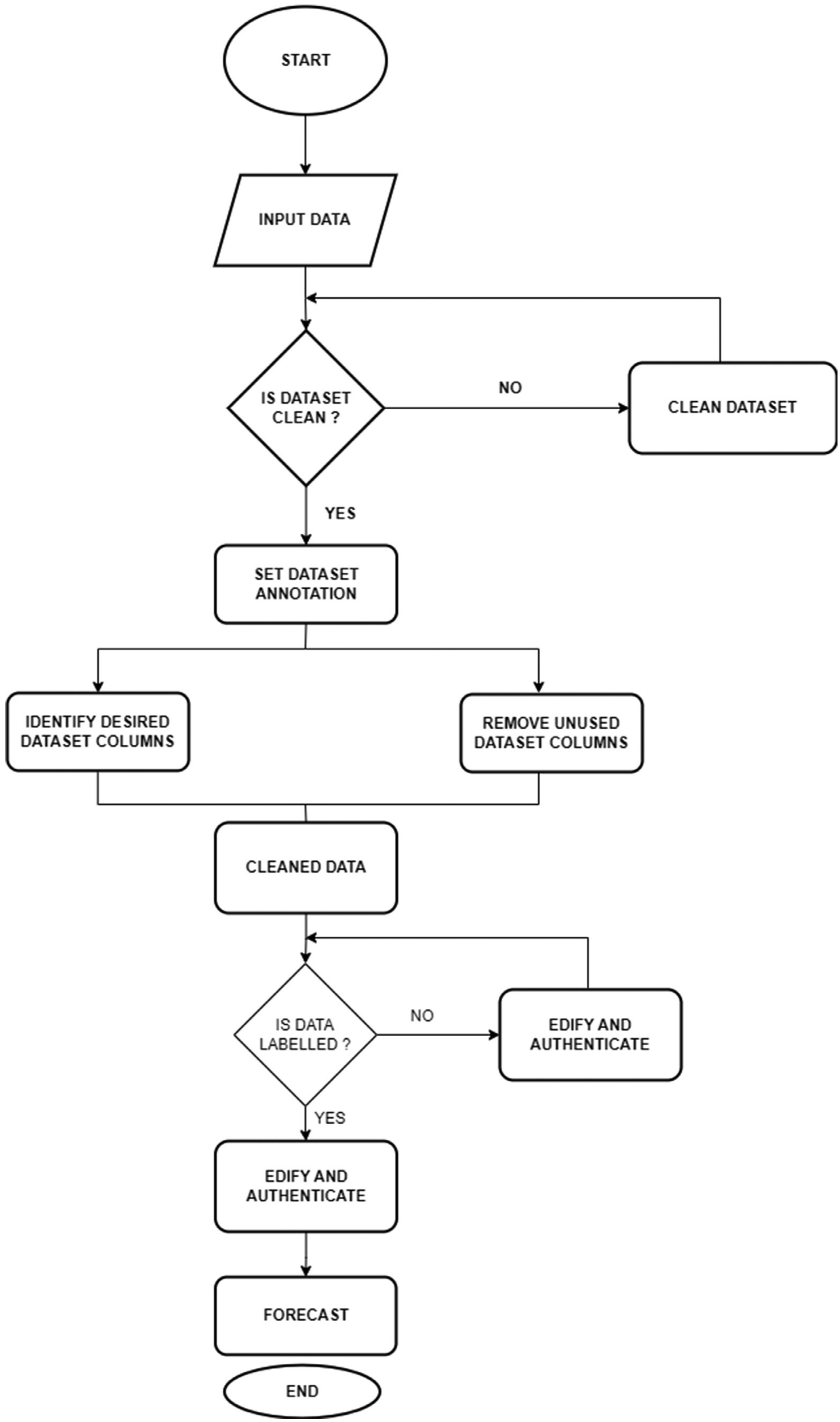

Fig. 13. Operational flow chart for proposed framework.

```
Begin
for affirmed Dataset
    If Dataset_notcleaned Do
    clean Dataset Do
        set Dataset_annotation
            for desired Dataset_columns
            Do remove unused Dataset_columns
            return extracted columns (Dataset)
            while Edify and Authenticate (Dataset)
    Else apply AssembleDeftAssessment (Dataset)
        if Dataset_cleaning_labelling_succesful Do
            edify_authenticate (Dataset)
    Foreach sentiment in dataset Do
        forecast (new prototype)
        //Measures and determines level of sarcastic statement
    ends
```

Fig. 14. Pseudocode for proposed framework.

Table 2
Tweeter dataset for sentiment analysis for aviation sector.

Class	Tweets
Non-sarcastic	9178 (62.7%)
Neutral	3099 (21.2%)
Sarcastic	2363 (16.1%)
Total	14,640 (100%)

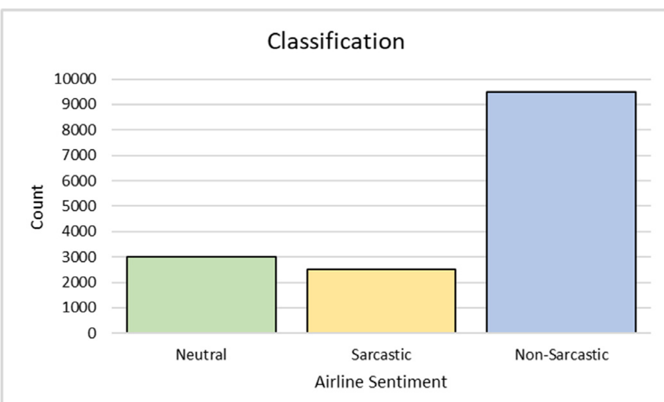

Fig. 15. Sentiment Classification.

We were able to authenticate and count the number of tweets in which the various airlines were referenced after analyzing them. The results are shown in Fig. 16. During the process, six airlines were counted.

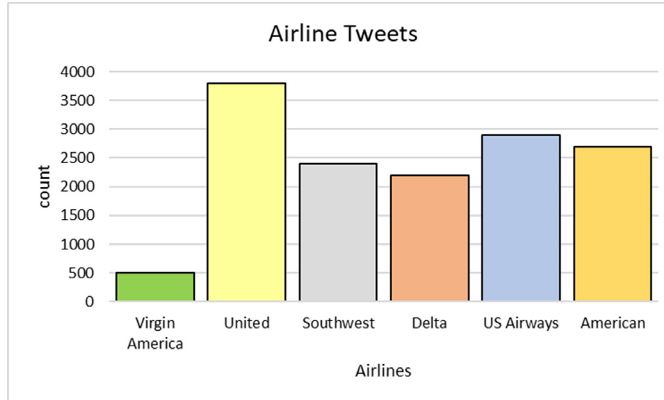

Fig. 16. Airline Tweet Count.

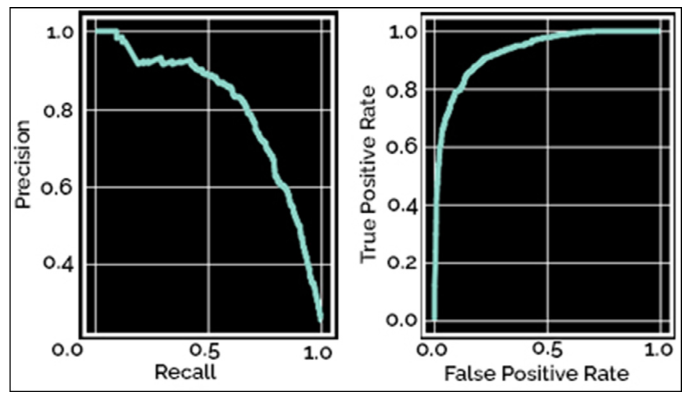

Fig. 17. Output of precision against recall and true - false positive rate.

Table 3
Results of standard SVM model with TF-IDF.

Train confusion matrix:		Test confusion matrix is:		
$\begin{bmatrix} 6824 & 31 \\ 151 & 1649 \end{bmatrix}$		$\begin{bmatrix} 2291 & 32 \\ 296 & 267 \end{bmatrix}$		
	Precision	Recall	F1-score	Support
0	0.89	0.99	0.93	2323
1	0.89	0.47	0.62	563
accuracy			0.89	2886
Macro avg	0.89	0.73	0.78	2886
Weighted avg	0.89	0.89	0.87	2886

Train accuracy score: 0.9789716926632005.
Test accuracy score: 0.886478863478863.
Train ROC-AUC Score: 0.9969059080962801.
Test ROC-AUC score: 0.929291531361801.
Area under Precision-Recall curve: 0.6194895591647333.
The area under ROC-AUC: 0.805035400076202.

4.1.1. *Text preparation*

The data is cleaned, and the important features are extracted as part of measures to clean and prepare the dataset for analysis as shown in Fig. 21.

4.1.2. *Base SVM model with TF-IDF*

A simple linear Support Vector Machine (SVM) classifier was created. Each unique word in the phrase, as well as all subsequent words, was considered by the classifier. We convert each word into a vector to make this format helpful for our SVM classifier. Our vocabulary, which is a list of all words discovered in our training data, has the same size as the vector, with each word representing an item in the vector. If a word appears in the vector, it has a value of 1; otherwise, it has a value of 0. To create these vectors, we use the Count Vectorizer (which makes it easy for text data to be used directly in machine learning and deep learning) from sklearn. For the SVM, a TF-IDF object is constructed. The training dataset is used to assess parameters such as precision, accuracy, recall, F1-score, and support (Fig. 17 and Table 3).

4.1.3. *Hyperparameter optimization to improve SVM sentiment analysis*

The results acquired after running our initial standard base SVM model with TF-IDF did not achieve optimal values, and hence will not provide us with the expected results. With the addition of hyperparameters, these outcomes can be improved. There are various hyperparameter optimization algorithms now available, but the Bayesian optimization strategy appears to be the most promising. The following are the two most prominent approaches to probability:

- Frequentist Approach: Focuses on the probability of the data given the hypothesis

```
# Tuning the hyperparameters
parameters ={
    "C":[0.1,2,11],
    "kernel":['linear', 'rbf', 'sigmoid'],
    "gamma":['scale', 'auto']
}


svm_optimal = grid_search(svm.SVC(probability=True), parameters,x_train, y_train)
```

Fig. 18. Code fragment of tuning of hyperparameters.

Table 4
Final Results of Optimized SVM model.

Train confusion matrix:		Test confusion matrix is:		
$\begin{bmatrix} 6829 & 26 \\ 8 & 1792 \end{bmatrix}$		$\begin{bmatrix} 2276 & 47 \\ 246 & 317 \end{bmatrix}$		
	Precision	Recall	F1-score	Support
0	0.90	0.98	0.94	2323
1	0.87	0.56	0.68	563
accuracy			0.90	2886
Macro avg	0.89	0.77	0.81	2886
Weighted avg	0.89	0.90	0.89	2886

Train accuracy score: 0.9968716348931253.
Test accuracy score: 0.8982753984753985.
Train ROC-AUC Score: 0.9982762784666505.
Test ROC-AUC score: 0.9311503086365474.
Are under Precision-Recall curve: 0.6839266450916937.
Area under ROC-AUC: 0.8125412196751889.

- Bayesian Approach: Focuses on the probability of the hypothesis given the data. That means fixed data and hypotheses are random.

$$\begin{array}{lll} \text{Statistics} & \text{Frequentist} & \theta \to x : p(x|\theta) \\ & \text{Bayesian} & \theta \to x : p(x|\theta) \\ & \theta : \text{Cause} & x : \text{Result} \end{array} \tag{4}$$

During the implementation of the optimization model, the following parameters are used:

$$k\left(x, x^{1}\right) = e^{-||\gamma x - x^{1}||^{2}} \tag{5}$$

Reasons for using the Bayesian Optimization Strategy:

- Bayesian Optimization picks a prior belief and then searches the parameter space by enforcing and updating that prior belief during training.
- Bayesian let their prior beliefs influence their predictions, while frequentists do not.

Steps for implementing the Optimization Algorithm:

- Introduction of key points to be used for the machine learning process
- Using previously evaluated points, compute a posterior expectation of loss
- Sample the loss at a new point, that maximizes some utility of the expectation (the best regions to sample from)

Fig. 18 presents a code fragment of the hyperparameter fine tuning resulting in an optimal parameter for the SVM code implementation. Table 4 presents the evaluation performance result for the optimized SVM model.

After the successful implementation of the framework to perform the sentiment analysis, results from three of the state-of-the-art techniques showed drastic improvement in the performance (Table 5). In review, a standard SVM model with TF-IDF (term frequency-inverse document frequency) gives a precision level of 0.89 for both positive (1) and negative (0) tweets. It also gave us a train accuracy score of 0.97 and a test accuracy score of 0.88 respectively.

Table 5
Results from three classification methods using the dataset.

Model Name	Train Accuracy	Test Accuracy	Train ROC	Test ROC
MultinomialNB	0.850029	0.834026	0.956111	0.901301
SVM	0.978972	0.886348	0.996906	0.929168
SVM Optimized	0.996072	0.898475	0.998276	0.931148

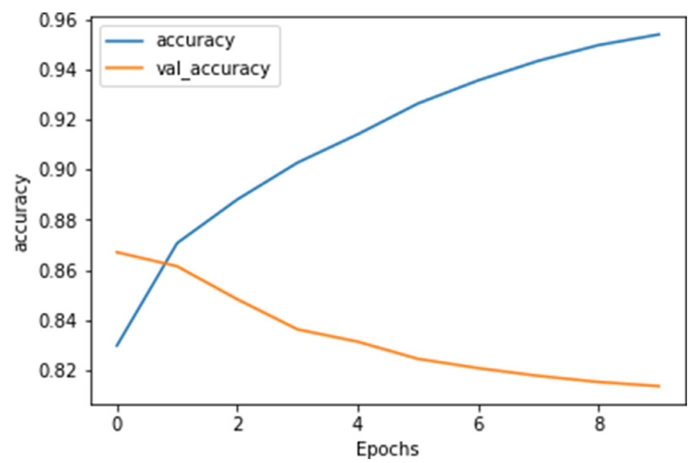

Fig. 19. Measure of Accuracy against Epochs.

However, after optimizing the hyperparameters with TF-IDF, the score improved to 0.90 for negative (0) tweets and 0.87 for positive tweets. The train accuracy scores also improved from 0.97 to 0.99, and the test accuracy score also improved from 0.88 to 0.89. This shows that using optimized hyperparameters with TF-IDF gives a better result than a standard SVM model.

In Fig. 19, an accuracy graph plot of GRU against epochs was plotted using the dataset. The accuracy level increased from 0.83 with epochs of 0 to 0.95 at an epoch of 8.5. The validation accuracy also decreased from 0.87 at 0 to 0.82 at 9.2. The GRU model achieved an accuracy of 84.94% with a loss score of 0.342 on the testing data. Fig. 20 presents the confusion matrix details with respective to the values for true positive, true negative, false positive and false negative.

5. Discussion

The main aim of this study is to introduce a framework comprising of three operators to classify opinion polls into sarcastic and non-sarcastic sentiments within the aviation industry. This study contributes significantly to a pool of knowledge in the field of information management

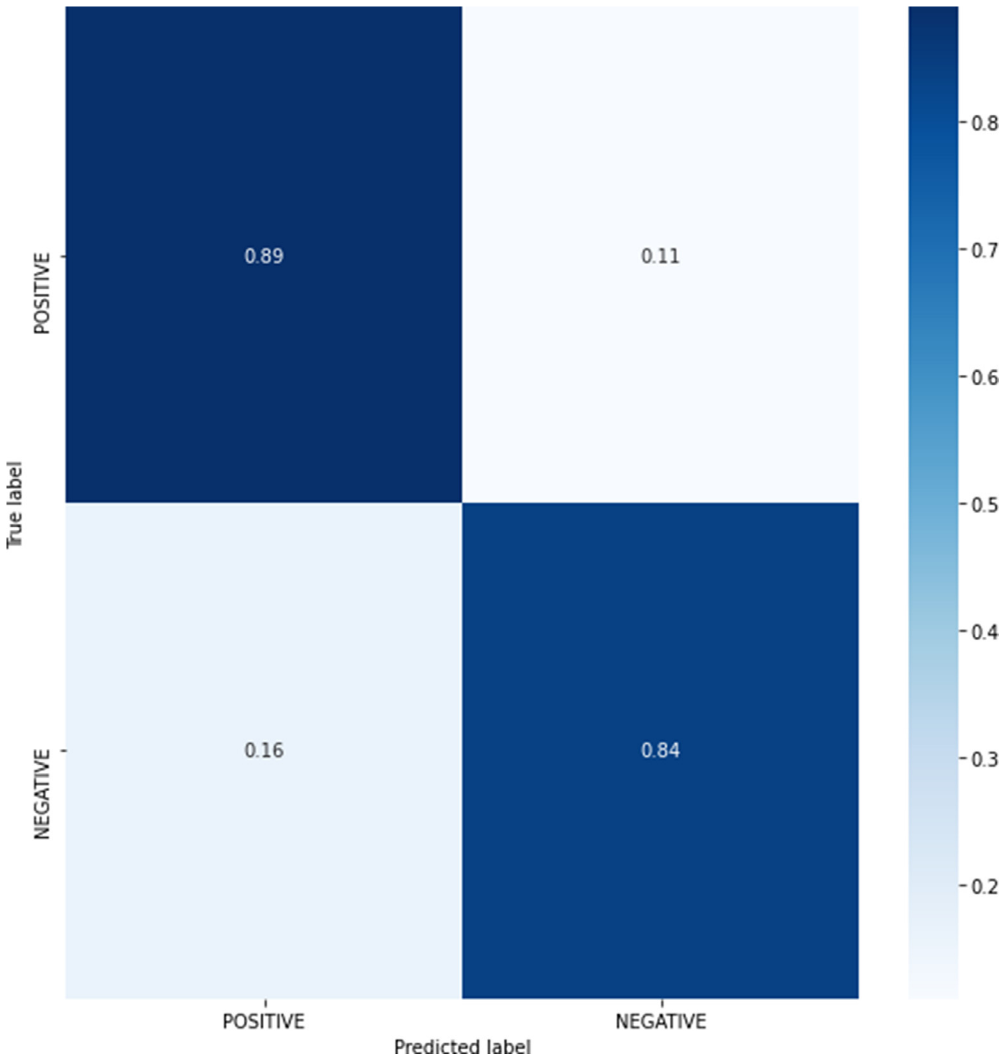

Fig. 20. Confusion matrix of GRU.

1	tweet_id	airline_sentiment	airline_sentiment_confidence	negativereason	negativereason_confidence	airline	name	text
2	0	neutral	1			Virgin America	cairdin	@VirginAmerica What @dhepburn said.
3	1	positive	0.3486		0	Virgin America	jnardino	@VirginAmerica plus you've added commercials to the experience... tacky.
4	2	neutral	0.6837			Virgin America	yvonnalynn	@VirginAmerica I didn't today... Must mean I need to take another trip!
5	3	negative	1	Bad Flight	0.7033	Virgin America	jnardino	@VirginAmerica it's really aggressive to blast obnoxious "entertainment" in your guests' fa
6	4	negative	1	Can't Tell	1	Virgin America	jnardino	@VirginAmerica and it's a really big bad thing about it
7	5	negative	1	Can't Tell	0.6842	Virgin America	jnardino	@VirginAmerica seriously would pay $30 a flight for seats that didn't have this playing.
8	6	positive	0.6745		0	Virgin America	cjmcginnis	@VirginAmerica yes, nearly every time I fly VX this â€œear wormâ€ wonâ€™t go away :)
9	7	neutral	0.634			Virgin America	pilot	@VirginAmerica Really missed a prime opportunity for Men Without Hats parody, there. htt
10	8	positive	0.6559			Virgin America	dhepburn	@virginamerica Well, I didn'tâ€¦but NOW I DO! :-D
11	9	positive	1			Virgin America	YupitsTate	@VirginAmerica it was amazing, and arrived an hour early. You're too good to me.
12	10	neutral	0.6769		0	Virgin America	idk_but_youtube	@VirginAmerica did you know that suicide is the second leading cause of death among teen
13	11	positive	1			Virgin America	HyperCamiLax	@VirginAmerica I <3 pretty graphics. so much better than minimal iconography. :D
14	12	positive	1			Virgin America	HyperCamiLax	@VirginAmerica This is such a great deal! Already thinking about my 2nd trip to @Australia
15	13	positive	0.6451			Virgin America	mollanderson	@VirginAmerica @virginmedia I'm flying your #fabulous #Seductive skies again! U take all t
16	14	positive	1			Virgin America	sjespers	@VirginAmerica Thanks!
17	15	negative	0.6842	Late Flight	0.3684	Virgin America	smartwatermelon	@VirginAmerica SFO-PDX schedule is still MIA.
18	16	positive	1			Virgin America	ItzBrianHunty	@VirginAmerica So excited for my first cross country flight LAX to MCO I've heard nothing b
19	17	negative	1	Bad Flight	1	Virgin America	heatherovieda	@VirginAmerica I flew from NYC to SFO last week and couldn't fully sit in my seat due to tw
20	18	positive	1			Virgin America	thebrandiray	I â¤ï¸ flying @VirginAmerica. â˜ºï¸ðŸ‘
21	19	positive	1			Virgin America	JNLpierce	@VirginAmerica you know what would be amazingly awesome? BOS-FLL PLEASE!!!!!!! I want
22	20	negative	0.6705	Can't Tell	0.3614	Virgin America	MISSGJ	@VirginAmerica why are your first fares in May over three times more than other carriers w
23	21	positive	1			Virgin America	DT_Les	@VirginAmerica I love this graphic. http://t.co/UT5GrRwAaA
24	22	positive	1			Virgin America	ElvinaBeck	@VirginAmerica I love the hipster innovation. You are a feel good brand.
25	23	neutral	1			Virgin America	rjlynch21086	@VirginAmerica will you be making BOS>LAS non stop permanently anytime soon?
26	24	negative	1	Customer Service Issu	0.3557	Virgin America	ayeevickiee	@VirginAmerica you guys messed up my seating.. I reserved seating with my friends and you
27	25	negative	1	Customer Service Issu	1	Virgin America	Leora13	@VirginAmerica status match program. I applied and it's been three weeks. Called and ema
28	26	negative	1	Can't Tell	0.6614	Virgin America	meredithjlynn	@VirginAmerica What happened 2 ur vegan food options?! At least say on ur site so i know
29	27	neutral	0.6854			Virgin America	AdamSinger	@VirginAmerica do you miss me? Don't worry we'll be together very soon.
30	28	negative	1	Bad Flight	1	Virgin America	blackjackpro911	@VirginAmerica amazing to me that we can't get any cold air from the vents. #VX358 #noair
31	29	neutral	0.615		0	Virgin America	TenantsUpstairs	@VirginAmerica LAX to EWR - Middle seat on a red eye. Such a noob maneuver. #sendambie

Fig. 21. Sample tweet data used for sentiment analysis.

through advancing the understanding of sentiment analysis, informing future research, and contributing to interdisciplinary research.

5.1. Contribution to literature

The literature contribution of this article is that it introduces a framework to assist the research community to detect and analyze opinion polls that will classify a given statement as sarcastic or non-sarcastic. Primarily, the framework makes use of three operators, namely *Assemble+Deft Assessment (ASSEMBLEDEFT), Edify & Authenticate,* and *Forecast of New Prototypes* to perform the sentiment classification analysis.

Despite the fact that a variety of frameworks have been developed in existing related works (Bharti et al., 2022; He et al., 2020; Khotijah et al., 2020; Xiong et al., 2019) and put to use to identify and classify sarcasm for sentiment analysis, the focus has not been on a more practical human-engaged services that is data-driven, for example in the aviation sector. The framework was tested using three state-of-the-art techniques, namely multinomialNB, SVM and SVM optimized (Carvalho et al., 2019; Kouadri et al., 2020; Swapnarekha et al., 2020), and it was observed that SVM optimized yielded improved classification performance. Thus, the use of SVM optimized is proposed as the best classification algorithm for sentiment analysis within the aviation sector.

This paper contributes to the information management literature through the following ways:

Advancing the understanding of sentiment analysis: This paper contributes to the advancement of sentiment analysis techniques by proposing a framework to classify instances of sarcastic sentiments within the aviation sector. This can assist researchers and practitioners better understand the nuances of sentiment analysis and improve the accuracy of sentiment analysis systems. An evidence-based study of the data creates value for open innovation through social media platforms and advances the usage of machine learning for such data driven paradigm (Adikari et al., 2021).

Enhancing information management in the aviation sector: The paper contributes to the enhancement of information management practices in the aviation sector by providing a framework to identify instances of sarcastic sentiment in customer feedback. Through user-generated content behavior, various amount of data on social media platforms becomes substantial to undertake analysis (Kitsios et al., 2022). This can assist airlines and aviation companies better understand customer feedback and take appropriate actions to improve their services.

Informing future research: The study provides a foundation for future research on sentiment analysis and information management in the aviation sector. Researchers can build on the proposed framework and further develop sentiment analysis techniques for the aviation industry.

Contributing to interdisciplinary research: This study contributes to the intersection of sentiment analysis, information management, and the aviation sector. This interdisciplinary approach can lead to new insights and perspectives on how sentiment analysis can be applied in different domains and industries.

5.2. Practical implications

The evolution and influence of social media has become a global phenomenon which has impacted the views and decisions of many aspects of our lives even at an individual level. With its explosive growth, the digital era provides organizations the platform to engage directly with their customers or end-users to better understand their needs (Grover et al., 2022). Engaging consumers on social media brand communities has also become imperative as it enables organizations to define the attitude of their customers which leads to further key findings (Santos et al., 2022). This has led to the generation of huge amount of online data.

By making use of the application of big data analytics (BDA) and natural language processing (Kumar et al., 2021) in emerging management disciplines such as the aviation sector, this study provides a structured and result oriented pathway to impact positively on the decision making of various service provision sectors. In a study by Kushwaha et al. (2021) which provided a reference for future information systems (IS) scholars to perform deep-drive analysis on such management area, the dynamic capabilities of BDA effectively classifies sarcastic and non-sarcastic sentiments to pave way for wider applications.

Not limited only to the detection, classification, and analysis of sarcasm on COVID-19 within the aviation industry, organizations and businesses which collects and utilizes online review data, social media comments, and customer feedbacks can maximize the use of the framework developed to make informed decisions to better serve their consumers. These final decisions can be channeled toward the development of improved products and offering better services through understanding and connecting with the opinions of the public and addressing any other concerns expressed through social media channels.

6. Conclusion

After comprehensive analysis of the importance of sentiment analysis to both industry and academia, two state-of-the-art techniques were employed to analyze data from Twitter concerning COVID-19 to determine the impact of the pandemic on a global stage. Comparatively, a framework was developed to evaluate the accuracy and performance of selected state-of-the-art machine learning and sentiment analysis techniques. In obtaining the dataset that was used for this study, a comprehensive approach was used. The data obtained from Twitter was put through a series of processing to extract, clean and filter the relevant information. This was done using state-of-the-art data processing steps. This allowed us to use the correct data to train our framework to obtain the right results.

A large amount of data related to COVID-19 and airline travel from Twitter was used to test the performance of the framework to estimate the impact of COVID-19 within the aviation industry. By implementing the proposed framework consisting of three main operators, namely *Assemble+Deft Assessment (ASSEMBLEDEFT), Edify & Authenticate* and *Forecast of New Prototypes,* we found an improved performance in the sentiment analysis and classification. The parameters used include precision, accuracy, recall and F1-score. The outcome of the framework showed significant increase from 9.28% under a standard sentiment review to 10.1% optimized sentiment analysis. With the implementation of the framework, there was an improved performance of the sentiment techniques.

By virtue of improvement in the performance of sentiment analysis using the parameters mentioned above, it can be concluded that the proposed framework can serve as a benchmark or baseline for future sentiment analysis and classification. This can be used in association with other machine learning or deep learning methods. This framework can also be used beyond the current topic of measuring sarcastic and non-sarcastic sentiments in the aviation industry. It can be used in any service industry where customer complaints can be collected and analyzed. The framework can be developed into a management information system for effective decision making by stakeholders within the aviation industry. This will provide support on how to better understand passengers' needs and address their traveling issues effectively amid any pandemic.

Funding

None.

Ethics approval

Not applicable.

Consent to participate

All authors agreed to participate in this research.

Consent for publication

All authors have agreed to publish this article.

Data source

The dataset considered for this study is sourced from Kaggle and contains comments of passengers on basis of service provided by airlines. Refer to Section 3.1 for details of the studied dataset.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abid, F., Alam, M., Yasir, M., & Li, C. (2019). Sentiment analysis through recurrent variants latterly on convolutional neural network of twitter. *Future Generation Computer Systems, 95*, 292–308. 10.1016/J.FUTURE.2018.12.018.

Adikari, A., Burnett, D., Sedera, D., de Silva, D., & Alahakoon, D. (2021). Value co-creation for open innovation: an evidence-based study of the data driven paradigm of social media using machine learning. *International Journal of Information Management Data Insights, 1*(2), Article 100022. 10.1016/j.jjimei.2021.100022.

Zhao, B., Jiang, X., & Cao, J. (2020). Biostatistical analysis on covid-19. Annals of Medical & Surgical Case Reports: AMSCR-100045. GRF Publishers. 2(03.

Zucco, C., Liang, H., Di Fatta, G., & Cannataro, M. (2018). Explainable sentiment analysis with applications in medicine. In IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 1740–1747). IEEE.

R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja (2019). The impact of features extraction on the sentiment analysis. Procedia Computer Science, 152, 341–348. 10.1016/j.procs.2019.05.008.

Alamoodi, A. H., Zaidan, B. B., Zaidan, A. A., Albahri, O. S., Mohammed, K. I., Malik, R. Q., et al., (2021). Sentiment analysis and its applications in fighting COVID-19 and infectious diseases : a systematic review. *Expert Systems With Applications, 167*, Article 114155 October 2020. 10.1016/j.eswa.2020.114155.

Alharbi, A. S. M., & de Doncker, E. (2019). Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. *Cognitive Systems Research, 54*, 50–61. 10.1016/j.cogsys.2018.10.001.

Aloqaily, A., Al-Hassan, M., Salah, K., Elshqeirat, B., & Almashagbah, M. (2020). Sentiment analysis for Arabic tweets datasets: Lexicon-based and machine learning approaches. *Journal of Theoretical and Applied Information Technology, 98*(4), 612–623.

Avinash, M., & Sivasankar, E. (2019). A study of feature extraction techniques for sentiment analysis. *Advances in Intelligent Systems and Computing, 814*, 475–486. 10.1007/978-981-13-1501-5_41.

Bharti, S. K., Babu, K. S., & Jena, S. K. (2015). Parsing-based sarcasm sentiment recognition in Twitter data. In *Proceedings of the IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM 2015* (pp. 1373–1380). 10.1145/2808797.2808910.

Bharti, S. K., Gupta, R. K., Pathik, N., & Mishra, A. (2022). Sarcasm detection in news headlines using voted classification. In *Proceedings of the ACM international conference proceeding series* (pp. 208–212). 10.1145/3549206.3549245.

Bhavitha, B. K., Rodrigues, A. P., & Chiplunkar, N. N. (2017). Comparative study of machine learning techniques in sentimental analysis. In *Proceedings of the international conference on inventive communication and computational technologies, ICICCT 2017, ICICCT* (pp. 216–221). 10.1109/ICICCT.2017.7975191.

Britzolakis, A., Kondylakis, H., & Papadakis, N. (2020). A review on lexicon-based and machine learning political sentiment analysis using tweets. *International Journal of Semantic Computing, 14*(4), 517–563. 10.1142/S1793351X20300010.

Brusaferri, A., Matteucci, M., Spinelli, S., & Vitali, A. (2020). Learning behavioral models by recurrent neural networks with discrete latent representations with application to a flexible industrial conveyor. *Computers in Industry, 122*, Article 103263. 10.1016/j.compind.2020.103263.

Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (2017). Affective Computing and Sentiment Analysis. In E. Cambria, D. Das, S. Bandyopadhyay, & A. Feraco (Eds.), *A Practical Guide to Sentiment Analysis. Socio-Affective Computing, 5*. Cham: Springer 10.1007/978-3-319-55394-8_1.

Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R., da, P., Basto, J. P., et al., (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers and Industrial Engineering, 137*, Article 106024 August. 10.1016/j.cie.2019.106024.

Chan, S. W. K., & Chong, M. W. C. (2017). Sentiment analysis in financial texts. *Decision Support Systems, 94*, 53–64. 10.1016/j.dss.2016.10.006.

Chiarello, F., Bonaccorsi, A., & Fantoni, G. (2020). Technical sentiment analysis. measuring advantages and drawbacks of new products using social media. *Computers in Industry, 123*, Article 103299. 10.1016/j.compind.2020.103299.

Cortes, C., Vapnik, V., & Saitta, L. (1995). Support-vector networks editor. *Machine leaming*: 20. Kluwer Academic Publishers.

Da Silva, N. F. F., Hruschka, E. R., & Hruschka, E. R (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems, 66*, 170–179. 10.1016/j.dss.2014.07.003.

Das, D., & Clark, A. J. (2018a). Sarcasm detection on facebook: a supervised learning approach. In *Proceedings of the 20th international conference on multimodal interaction, ICMI 2018*. 10.1145/3281151.3281154.

Das, D., & Clark, A. J. (2018b). Sarcasm detection on Flickr using a CNN. In *Proceedings of the ACM international conference proceeding series* (pp. 56–61). 10.1145/3277104.3277118.

El Rahman, S. A., AlOtaibi, F. A., & AlShehri, W. A. (2019). *Sentiment Analysis of Twitter Data, 2019 International Conference on Computer and Information Sciences (ICCIS)* (pp. 1–4). Sakaka, Saudi Arabia. IEEE. 10.1109/ICCISci.2019.8716464.

Gkikas, D. C., Tzafilkou, K., Theodoridis, P. K., Garmpis, A., & Gkikas, M. C. (2022). How do text characteristics impact user engagement in social media posts: modeling content readability, length, and hashtags number in Facebook. *International Journal of Information Management Data Insights, 2*(1), Article 100067. 10.1016/j.jjimei.2022.100067.

Grover, P., Kar, A. K., & Dwivedi, Y. (2022). The evolution of social media influence - a literature review and research agenda. *International Journal of Information Management Data Insights, 2*(2), Article 100116. 10.1016/j.jjimei.2022.100116.

Hassan, A., & Mahmood, A. (2017). Deep Learning approach for sentiment analysis of short texts. In *Proceedings of the 3rd international conference on control, automation and robotics, ICCAR 2017* (pp. 705–710). 10.1109/ICCAR.2017.7942788.

He, S., Guo, F., & Qin, S. (2020). Sarcasm detection using graph convolutional networks with bidirectional LSTM. In *Proceedings of the ACM international conference proceeding series* (pp. 97–101). 10.1145/3422713.3422722.

Hemmatian, F., & Sohrabi, M. K. (2017). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review, 52*(3), 1495–1545 2017 52:3. 10.1007/S10462-017-9599-6.

Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences, 30*(4), 330–338. 10.1016/j.jksues.2016.04.002.

Jain, K., & Kaushal, S. (2018). A comparative study of machine learning and deep learning techniques for sentiment analysis. In 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 483–487). IEEE. 10.1109/ICRITO.2018.8748793.

Jangid, H., Singhal, S., Shah, R. R., & Zimmermann, R. (2018). Aspect-based financial sentiment analysis using deep learning. *ACM Digital Library*, 1961–1966. 10.1145/3184558.3191827.

Kandasamy, I., Vasantha, W. B., Obbineni, J. M., & Smarandache, F. (2020). Sentiment analysis of tweets using refined neutrosophic sets. *Computers in Industry, 115*, Article 103180. 10.1016/j.compind.2019.103180.

Kar, A. K., Choudhary, S. K., & Singh, V. K. (2022). How can artificial intelligence impact sustainability: a systematic literature review. *Journal of Cleaner Production, 376*, Article 134120 May. 10.1016/j.jclepro.2022.134120.

Kazmaier, J., & van Vuuren, J. H. (2020). A generic framework for sentiment analysis: leveraging opinion-bearing data to inform decision making. *Decision Support Systems, 135*, Article 113304 April. 10.1016/j.dss.2020.113304.

Khaleghparast, S., Maleki, M., Hajianfar, G., Soumari, E., Oveisi, M., Golandouz, H. M., et al., (2023). Development of a patients' satisfaction analysis system using machine learning and lexicon-based methods. *BMC Health Services Research, 23*(1), 280. 10.1186/s12913-023-09260-7.

Khotijah, S., Tirtawangsa, J., & Suryani, A. A. (2020). Using LSTM for context based approach of sarcasm detection in twitter. In *Proceedings of the ACM international conference proceeding series* (pp. 1–7). 10.1145/3406601.3406624.

Kitsios, F., Mitsopoulou, E., Moustaka, E., & Kamariotou, M. (2022). User-generated content behavior and digital tourism services: A SEM-neural network model for information trust in social networking sites. *International Journal of Information Management Data Insights, 2*(1), Article 100056. 10.1016/j.jjimei.2021.100056.

Kouadri, W. M., Ouziri, M., Benbernou, S., Echihabi, K., Palpanas, T., & Amor, I. B (2020). Quality of sentiment analysis tools: the reasons of inconsistency. In *Proceedings of the VLDB endowment: 14* (pp. 668–681). 10.14778/3436905.3436924.

Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: a systematic literature review. *International Journal of Information Management Data Insights, 1*(1), Article 100008. 10.1016/j.jjimei.2021.100008.

Kushwaha, A. K., Kar, A. K., & Dwivedi, Y. K. (2021). Applications of big data in emerging management disciplines: A literature review using text mining. *International Journal of Information Management Data Insights, 1*(2), Article 100017. 10.1016/j.jjimei.2021.100017.

Kusyanti, A., & Zakia, D. (2019). Sentiment Analysis in the Mobile Application Review Document Using the Improved K-Nearest Neighbor Method. In 2019 International Conference on Sustainable Information Engineering and Technology (SIET) (pp. 332–337). IEEE. 10.1109/SIET48054.2019.8986037.

Lee, J., Woo, J., Kang, A. R., Jeong, Y. S., Jung, W., Lee, M., et al., (2020). Comparative analysis on machine learning and deep learning to predict post-induction hypotension. *Sensors, 20*(16), 1–21 Switzerland. 10.3390/s20164575.

L. Li, T.T. Goh, and D. Jin (2018). How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis. Neural Computing and Applications 2018 32:9, 32(9), 4387–4415. 10.1007/S00521-018-3865-7.

Mehta, P., & Pandya, S. (2020). A review on sentiment analysis methodologies. *Practices and Applications, 9*(02), 601–609.

Morariu, C., Morariu, O., Răileanu, S., & Borangiu, T. (2020). Machine learning for predictive scheduling and resource allocation in large scale manufacturing systems. *Computers in Industry, 120*. 10.1016/j.compind.2020.103244.

Qian, J., Niu, Z., & Shi, C. (2018). Sentiment analysis model on weather related tweets with deep neural network. In *Proceedings of the ACM international conference proceeding series* (pp. 31–35). 10.1145/3195106.3195111.

Rajadesingan, A., Zafarani, R., & Liu, H. (2015). Sarcasm detection on twitter:a behavioral modeling approach. In *WSDM - Proceedings of the 8th ACM international conference on web search and data mining* (pp. 97–106). 10.1145/2684822.2685316.

Ri, H., Dqg, F., Ri, U., Pfv, Q. S., Qg, N. H. G. X., Ri, H., et al., (2021). Techniques of sarcasm detection: a review. In *Proceedings of the international conference on advance computing and innovative technologies in engineering: 7* (pp. 5–9).

Santos, Z. R., Cheung, C., Coelho, P. S., & Rita, P. (2022). Consumer engagement in social media brand communities: a literature review. *International Journal of Information Management, 63* November 2021. 10.1016/j.ijinfomgt.2021.102457.

Satapathy, R., Cambria, E., & Hussain, A. (2017). Sentiment Analysis in the Bio-Medical Domain (p. 7). ISBN : 978-3-319-68467-3.

Saura, J. R., Palos-Sanchez, P., Angel, M., & Martin, R. (2018). Attitudes expressed in online comments about environmental factors in the tourism sector: an exploratory study. *International Journal of Environmental Research and Public Health Article*. 10.3390/ijerph15030553.

S. Sohangir, D. Wang, A. Pomeranets, and T.M. Khoshgoftaar (n.d.). Big data: deep learning for financial sentiment analysis. Journal of Big Data. 10.1186/s40537-017-0111-6.

Sufi, F. K. (2022). Identifying the drivers of negative news with sentiment, entity and regression analysis. *International Journal of Information Management Data Insights, 2*(1), Article 100074. 10.1016/j.jjimei.2022.100074.

Swapnarekha, H., Behera, H. S., Nayak, J., & Naik, B. (2020). Role of intelligent computing in COVID-19 prognosis: a state-of-the-art review. *Chaos, Solitons & Fractals, 138*, Article 109947. 10.1016/J.CHAOS.2020.109947.

Joachims, T. (2019). Text categorization with support vector machines: learning with many relevant features. *Culture Shift in Advanced Industrial Society*, 335–370. 10.1515/9780691186740-014.

Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. *Tropical Medicine & International Health, 25*(3), 278–280. 10.1111/TMI.13383.

Wang, P., Fan, E., & Wang, P. (2021). Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognition Letters, 141*, 61–67. 10.1016/j.patrec.2020.07.042.

Xiong, T., Zhang, P., Zhu, H., & Yang, Y. (2019). Sarcasm detection with self-matching networks and low-rank bilinear pooling. In *Proceedings of the web conference - proceedings of the world wide web conference, WWW 2019* (pp. 2115–2124). 10.1145/3308558.3313735.

Yakubu, H., & Kwong, C. K. (2021). Technological forecasting & social change forecasting the importance of product attributes using online customer reviews and google trends. *Technological Forecasting & Social Change, 171*, Article 120983 June. 10.1016/j.techfore.2021.120983.

H. Yakubu and C. K. Kwong, Using Online Big Data for Determining the Importance of Product Attributes, 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, Singapore, 2020, pp. 691–695, 10.1109/IEEM45057.2020.9309746.