Analyzing Public Sentiment Towards the Covid-19 Pandemic: A Twitter-based Sentiment Analysis and Machine Learning Approach

Andreas Kanavos*, Nikos Antonopoulos†, Alaa Mohasseb‡ and Phivos Mylonas§
*Department of Informatics, Ionian University, Corfu, Greece
akanavos@ionio.gr
†Department of Digital Media and Communication, NeMeCULAB, Ionian University, Argostoli, Greece
nikos@antonopoulos.info
‡School of Computing, University of Portsmouth, Portsmouth, UK
alaa.mohasseb@port.ac.uk
§Department of Informatics and Computer Engineering, University of West Attica, Athens, Greece
mylonasf@uniwa.gr

2023 18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP) | 979-8-3503-2771-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/SMAP59435.2023.10255176

***Abstract*—The Covid-19 pandemic has significantly reshaped societies, prompting an unprecedented surge in digital interactions and communication via social media platforms. Amidst this evolving landscape, the analysis of public sentiment towards pandemic management strategies has emerged as a critical avenue for understanding societal responses. This paper delves into the intricate landscape of sentiment dynamics surrounding the Covid-19 pandemic, with a focus on sentiments expressed in Twitter posts. Leveraging sentiment analysis techniques, the study provides nuanced insights into societal attitudes, concerns, and perceptions across distinct phases of the pandemic. The investigation not only employs traditional lexicon-based sentiment analysis but also explores the intersection of sentiment analysis and machine learning algorithms. The research contributes to the discourse by presenting a comprehensive analysis of public sentiment dynamics during various pandemic stages, shedding light on evolving emotional responses and offering insights into the effectiveness of measures.**

***Index Terms*—Sentiment Analysis, Covid-19, Social Media, Twitter, Machine Learning, Natural Language Processing, Opinion Mining, Text Mining**

I. Introduction

The Covid-19 pandemic has presented an unparalleled challenge to societies worldwide, fundamentally reshaping norms, values, and daily routines. Amidst this backdrop, the significance of public sentiment and its analysis has surged, becoming a focal point for understanding how individuals perceive and respond to the evolving crisis [18], [28]. As social interactions and communication increasingly pivot to digital platforms, particularly social media, these platforms have evolved into virtual arenas for expressing emotions, concerns, and opinions regarding the pandemic and its management strategies.

Twitter, as a prominent social networking platform, stands at the forefront of this transformation [10]. Its real-time nature and concise message format make it an ideal medium for individuals to share their thoughts, emotions, and reactions to ongoing events, such as the Covid-19 pandemic [20], [21]. Users turn to Twitter not only to express their sentiments but also to engage with others, access information, and seek solace in shared experiences [9]. Consequently, analyzing sentiments expressed in Twitter posts has emerged as a potent method to gain insights into the collective mindset of a society grappling with a crisis [1], [5], [30].

In this context, this research endeavors to explore the intricate landscape of public sentiment towards the management of the Covid-19 pandemic. By analyzing sentiments expressed in social media content, particularly Twitter posts, the study aims to unearth valuable insights into societal attitudes, concerns, and perceptions during various critical periods of the crisis.

This paper contributes to the ongoing discourse by presenting a comprehensive analysis of public sentiment surrounding the Covid-19 pandemic. By harnessing the power of sentiment analysis techniques, the study investigates the sentiments manifested in a vast collection of Twitter posts, offering multifaceted insights into the ebb and flow of emotions, concerns, and opinions during distinct phases of the pandemic. Furthermore, the research delves into the domain of supervised machine learning, employing a range of algorithms and text encodings to unveil nuanced sentiment trends. Through this approach, the paper illuminates how the fusion of NLP and machine learning can deepen our understanding of public sentiment dynamics.

The subsequent structure of the paper unfolds as follows: Section II offers an overview of existing research in the field, accentuating the novelty and contributions of this study. Section IV elucidates the core concepts and methodologies employed in the investigation. In Section V, research findings are expounded, along with insights into the utilized dataset. Finally, Section VI concludes the paper by summarizing key insights and delineating potential avenues for future research.

II. Related Work

Social network analysis of tweets related to the Covid-19 pandemic using machine learning techniques has become a prevalent domain within data mining, drawing on an extensive and continually expanding body of literature. In an early endeavor [31], authors employed Latent Dirichlet Allocation (LDA) to discern unigrams, bigrams, significant topics, themes, and sentiments from a dataset of four million Twitter messages spanning March 1 to April 21, 2020, concerning Covid-19. Another LDA-based work [13] tackled topic modeling, centering on identifying prevailing sentiments amid the pandemic, with fear emerging as the dominant emotion.

Furthermore, a study by [2] conducted social network and content analysis on tweets collected between March 27 and April 4, 2020, to investigate the conspiracy theory linking 5G towers to the pandemic's spread. Ensembles of classifiers were explored for tweet sentiment analysis [8], comparing strategies such as bag-of-words and feature hashing for tweet representation. The examination of a database of Covid-19-related tweets aiming to shed light on mask usage [23] involved classification to categorize tweets into high-level themes and subtopics within each theme.

LSTM neural networks have gained prominence for forecasting Covid-19 infection rates across countries, particularly during lockdown periods [7], [27], [28]. In a study analyzing posts from the Chinese social media platform Sina Weibo [29], a fine-tuned unsupervised BERT model and Tf-Idf model were employed for sentiment classification, with posts classified as Negative, Neutral, or Positive. Sentiment classification was also addressed in [22], utilizing Naive Bayes and Logistic Regression techniques on datasets of varying tweet lengths, revealing superior accuracy in the former method.

III. Preliminaries

A. Sentiment Analysis

Sentiment Analysis, also known as Opinion Mining, is a field that involves automated techniques applied to text in order to uncover the opinions expressed in relation to a specific topic or entity. This could range from products, films, books, public figures, groups, political parties or candidates, events, situations, or societal issues. Sentiment Analysis plays a crucial role in understanding and utilizing the abundant online data available today, aiming to provide insights into the polarity (emotional orientation), emotional states (such as joy, sadness, anger), intentions, and subjectivity or objectivity expressed in a text [16], [17].

In this work, our focus will primarily be on sentiment analysis methods related to polarity categorization. The main objective of sentiment analysis is to determine the emotional orientation of a text, whether it's positive, negative, or neutral.

Approaches to sentiment analysis can be broadly categorized into semantic orientation techniques and machine learning techniques, with or without supervision. In semantic orientation, a prevalent approach involves the use of sentiment dictionaries, which are pre-built resources containing words and expressions matched with specific emotions. On the other hand, the ML approach utilizes advanced NLP techniques and learning algorithms to train models capable of recognizing emotions.

A key distinction between these approaches lies in their independence from the subject matter of the analyzed texts. Sentiment analysis models based on Dictionaries tend to offer more stable performance, relying heavily on the quality of the employed Dictionary rather than the content of the analyzed texts. Conversely, ML-based sentiment analysis models heavily rely on the relevance of the subject matter in the analyzed texts and the texts used for training the model [26].

B. Lexicon-based Sentiment Analysis

Lexicon-based Sentiment Analysis is a widely used approach that falls under the category of Semantic Orientation techniques [14]. This method is simple yet effective and doesn't require extensive data pre-processing. It relies on pre-existing emotion dictionaries, which contain words and expressions associated with specific emotional orientations. In this approach, emotions in a text are identified based on the frequency of emotionally charged words and expressions.

Dictionaries predominantly focus on adjectives and adverbs, which characterize entities and situations in a sentence. Some dictionaries also include nouns, verbs, and expressions, and specialized dictionaries might be used depending on the analysis's topic. Some dictionaries go beyond binary polarity and assign an emotional value to each word within a predefined emotional rating scale. In a simplified process, assuming a single expressed opinion (positive, negative, or neutral) in the analyzed text:

- The text is segmented into sentences.
- Each sentence is further divided into words or expressions.
- Each word or expression is checked against the emotion dictionary.
- The frequency and associated emotion of each matched word or expression are recorded.
- Emotion counters are updated to identify the sentence's dominant emotion and estimate the sentiment.

This streamlined process captures the essence of using dictionaries for sentiment analysis. Variations exist based on specific implementations.

Lexicon-based sentiment analysis offers advantages like manual curation by humans for accurate emotional capture, reusability, and the potential for enrichment with synonyms or antonyms. It remains relatively stable in performance across various fields or topics.

On the other hand, the dictionary sentiment analysis approach operates under the assumption that a text's polarity can be inferred by calculating the polarity of its individual words. However, due to the intricate nature of natural language, this straightforward approach often falls short, as it fails to consider various linguistic nuances, including factors like the presence of negation [19].

Several widely recognized vocabularies are utilized in sentiment analysis:

1) SentiWordNet: SentiWordNet is a sentiment-focused dictionary developed specifically for knowledge mining applications and is based on the WordNet dictionary's structure. WordNet is an interpretive dictionary of the English language, categorizing nouns, verbs, adjectives, and adverbs into sets of cognitive synonyms known as synsets. These synsets express distinct concepts and are interconnected through conceptual-semantic and lexical relations.

In SentiWordNet, each term in WordNet is associated with a vector of three parameters: Pos(s), Neg(s), and Obj(s). These parameters, obtained through automated machine learning techniques including weak supervision, semi-supervised learning, and random-walk steps, describe the intensity of positive, negative, or neutral sentiment evoked by the dictionary term. The values of these parameters range from 0.0 to 1.0, ensuring that their sum equals 1 [3].

2) Afinn: Afinn is a widely used sentiment analysis dictionary, particularly popular in NLP and Twitter sentiment analysis. The dictionary is manually created and available in a text file format. Afinn assigns English words and certain phrases frequently used in microblogging with integer values in the range (-5.5) to indicate sentiment polarity scores. The vocabulary comprises 2,477 English words and selected microblogging phrases.

3) SenticNet: SenticNet introduces an alternative approach to sentiment analysis based on the field of Sentic Computing [6]. This vocabulary operates at the concept level, where sentiment recognition is not solely dependent on the frequency of word and expression recurrence, but also incorporates their expanded meanings.

Unlike other dictionaries, SenticNet is not built on any pre-existing semantic dictionary. Instead, it was generated automatically by applying Graph Mining and Multidimensional Scaling techniques to common-sense data from repositories such as WordNet-Affect, OMCS (Open Mind Common Sense), and GECKA (Game Engine for Common Sense Knowledge Acquisition). These techniques provide semantics, sentiments, and polarity related to approximately 200,000 natural language concepts. SenticNet employs a sentiment scale within the interval of (-1,1) to determine polarity.

4) TextBlob: TextBlob is a widely used and highly regarded open-source library for NLP and sentiment analysis tasks [25]. It offers a comprehensive set of tools for processing textual data, including sentiment analysis. One of TextBlob's strengths lies in its robust sentiment dictionary, which enhances its efficiency in analyzing text sentiment. The dictionary is enriched with semantic tags, which are utilized both internally in TextBlob's processes and are also available in XML file format for users' reference.

TextBlob employs a sentiment scale within the interval of (-1,1) to determine the polarity of text. This scale allows for a nuanced evaluation of sentiment, accommodating both positive and negative orientations.

IV. METHODOLOGY

To achieve the objectives of this study, a systematic methodology is employed to analyze sentiment towards the Covid-19 pandemic using Twitter data. The methodology comprises the following steps:

A. Data Collection

The initial step involves collecting a substantial volume of Twitter posts (tweets) related to Covid-19 and its mutations. The Twitter API is accessed through Python programming to retrieve real-time tweets meeting specific search criteria. These criteria encompass timeframes, and specialized keywords to ensure pertinent data collection [11], [12].

B. Data Pre-processing

Upon collection, the acquired tweets undergo a comprehensive pre-processing phase aimed at priming them for subsequent sentiment analysis and machine learning procedures. This pre-processing stage is vital to ensure that the subsequent algorithms are fed clean and cohesive textual input.

The pre-processing pipeline encompasses a series of actions, meticulously designed to eliminate extraneous elements that contribute minimal to no semantic value to the text [4]. The following actions are performed:

- Removal of URLs, numbers, symbols, punctuation marks, personal references (@username), hashtags (#), emojis, and trailing spaces.
- Conversion of text to lowercase characters to ensure uniformity and avoid discrepancies.
- Elimination of duplicate tweets to streamline subsequent analyses and avoid data redundancy.
- Removal of stopwords to focus exclusively on words carrying substantive meaning.

The pre-processed tweets, now stripped of superfluous components, are then stored for further analysis. The cleansing process applied to the training data mirrors that of the retrieved tweets prior to their entry into sentiment analysis methods with dictionaries.

Subsequently, both the training data and the retrieved tweets undergo an additional pre-processing stage encompassing lemmatization and stemming tasks. These tasks serve to curtail the dimensions of the vectors generated during the subsequent text encoding phase. This process optimizes the efficiency of vectorization for later stages of analysis.

C. Sentiment Analysis with Dictionaries

The sentiment analysis commences using dictionary-based approaches, employing Afinn, SenticNet, SentiWordNet, and TextBlob dictionaries.

D. Training Data Acquisition

For training machine learning models, a set of tagged tweets is essential. These tweets are labeled based on the emotions they convey. The labeled data is used for training and evaluating the predictive models, with an 80%-20% split for training and evaluation respectively.

E. Text Encoding and Vectorization

Further pre-processing steps are implemented, including lemmatization and stemming, followed by encoding the cleaned tweet data into vector representations. This vectorization process enhances the efficiency of subsequent machine learning techniques. The encoded data is used to facilitate different encoding strategies (N-Grams) for comparison.

F. ML-based Sentiment Analysis

The sentiment analysis extends to ML techniques. Logistic Regression, Multilayer Perceptron (MLP) Neural Network, and Naive Bayes algorithms for sentiment classification were implemented. These algorithms are evaluated using the encoded tweet data, considering different N-Gram encoding approaches.

V. Experimental Evaluation

A. Dataset

The experimental evaluation is grounded in a dataset of Twitter posts (tweets) sourced globally to capture a comprehensive view of sentiments surrounding the Covid-19 pandemic and its mutations. To ensure thematic relevance, a set of retrieval terms (keywords) was curated. These terms encompass phrases and hashtags linked to the pandemic's evolution and variants like Delta and Epsilon mutations. Examples include "#deltavariant," "#epsilonvariant," and "#Covid19."

This curated set of keywords ensured the collection of a representative sample of tweets. A total of 30,000 tweets were collected from July 11th to 14th, 2022.

Post pre-processing, the dataset reduced to 23,039 unique and cleansed tweets. The cleaning process prepares data for sentiment analysis and enhances analysis accuracy. Following the lemmatization and stemming stage, a representative sample of the retrieved tweets is showcased. Notably, at this juncture, the total count of tweets earmarked for analysis using ML techniques has been streamlined to 22,954 unique tweets.

B. Training Dataset

To expedite training machine learning models, a pre-existing dataset[1] containing 1,600,000 tagged tweets was procured. From this, 250,000 tweets were extracted for specific model training.

This subset forms the cornerstone for predictive model training and evaluation, with an 80%-20% training-evaluation ratio. Training data tweets have pre-assigned emotional labels, obtained using an emoticon detection algorithm.

Each entry within the training dataset features the following fields:

- Sentiment: Label (0 for negative, 4 for positive).
- ID: Unique identifier.
- Date: Publication date.
- Username: Author's username.
- Body: Textual content.

[1]https://www.kaggle.com/datasets/kazanova/sentiment140

Prior to training, the data undergoes pre-processing mirroring that of retrieved tweets. Lemmatization and stemming tasks follow, reducing vector dimensions for efficient text encoding. The 250,000 tweets are reduced to 213,876 after this process, with 80% (171,100 tweets) used for training.

C. Data Analysis

The collected tweets underwent comprehensive analysis to extract insights on public sentiment towards the pandemic. Dictionary-based approach used Afinn, SenticNet, SentiWordNet, and TextBlob dictionaries. Machine learning approach utilized Logistic Regression, MLP Neural Network and Naive Bayes algorithms. Both approaches were evaluated using different encoding strategies (N-Grams).

Subsequently, a new dataset comprising 236,830 tweets is created as shown in Table I.

TABLE I
DATASET INFORMATION

Details	Value
Period Examined	11-14/7/2022
Number of Retrieved Tweets	30,000
Total Tweets Used	22,954
Number of Training Dataset Tweets	250,000
Total Training Tweets Used	213,876
Tweets for ML Model Training	171,100
Tweets for ML Model Evaluation	42,776

D. Sentiment Analysis Results with Dictionaries

In the context of dictionary-based sentiment analysis from Twitter users' posts, four distinct dictionary approaches were employed: Afinn, SenticNet, SentiWordNet, and TextBlob as depicted in Table II.

The analysis using TextBlob dictionary reveals that the distribution of sentiments is as follows: neutral emotion dominates at a rate of 47.09%, with a substantial percentage of 33.96% expressing positive sentiments towards the response and progress of the pandemic. Notably, only 18.95% of the analyzed tweets convey strong concern about the evolution of Covid-19 and its mutations. For the Afinn dictionary, the results show that neutral sentiment prevails at a rate of 45.93%, while the remaining 54.07% is almost evenly split between positive and negative sentiments, with a slightly higher percentage of negative emotions expressed through tweets. With SentiWordNet dictionary, positive sentiments of optimism dominate at a rate of 42.81%, while notably, a significant portion of 34.62% is expressed negatively through tweets, underscoring the coexistence of two contrasting emotions. The SenticNet dictionary reveals that the positive sentiment of optimism prevails with a percentage approaching 57.61%. Similar to the SentiWordNet analysis, a considerable portion of tweets, around 31.73%, expresses negative emotions associated with concerns about the pandemic's evolution.

Comparing the dictionaries, TextBlob and Afinn, both of which employ simplified parsing procedures and a limited vocabulary range, yield similar results, with the neutral emotion

TABLE II
SENTIMENT ANALYSIS USING DICTIONARIES

	Positive Tweets	Positive Tweets (%)	Neutral Tweets	Neutral Tweets (%)	Negative Tweets	Negative Tweets (%)
Afinn	5,931	25.74%	10,582	45.93%	6,526	28.33%
SenticNet	13,273	57.61%	2,456	10.66%	7,310	31.73%
SentiWordNet	9,862	42.81%	5,200	22.57%	7,977	34.62%
TextBlob	7,823	33.96%	10,849	47.09%	4,367	18.95%

prevailing at over 45%. In contrast, the dictionaries SentiWordNet and SenticNet, utilizing more sophisticated analysis methods and a broader vocabulary, offer consistent outcomes. They reveal the dominance of two contrasting emotions: a positive sentiment of optimism, surpassing 42%, and a negative sentiment of anxiety about the pandemic's evolution, exceeding 31%.

E. Sentiment Analysis Results with ML Methods

For supervised ML sentiment analysis from Twitter users' posts, three different algorithmic approaches were implemented: Logistic Regression, MLP Neural Network and Naive Bayes. These algorithms were evaluated using four different text encoding N-Grams (1-Gram, 2-Grams, 3-Grams, and 4-Grams). The performance of the implemented supervised ML algorithms for each N-Gram encoding was examined. Table III presents the results of the evaluation of prediction models calculated for each N-Gram encoding.

TABLE III
PERFORMANCE OF ML ALGORITHMS WITH N-GRAM ENCODINGS

Algorithm	1-Grams	2-Grams	3-Grams	4-Grams
Logistic Regression	0.739	0.733	0.710	0.676
MLP Neural Network	0.737	0.732	0.709	0.661
Naive Bayes	0.728	0.718	0.693	0.655

We observe that the 1-Grams encoding is the most efficient for all tested algorithms, followed closely by the 2-Grams encoding. The top three algorithm-encoding combinations in terms of efficiency were isolated for further examination:

- Logistic Regression with 1-Grams encoding achieved the highest performance at 73.9% accuracy.
- MLP Neural Network with 1-Grams encoding demonstrated a minimal performance difference of 73.7%.
- Logistic Regression with 2-Grams encoding secured the third place with a yield of 73.3%.

In order to facilitate a concise comparison of the three selected algorithm-encoding combinations, summary Table IV follows.

TABLE IV
SENTIMENT ANALYSIS WITH MACHINE LEARNING METHODS

Algorithm	Encoding	Accuracy	Pos (%)	Neg (%)
Logistic Regression	1-Grams	73.9%	57.6%	42.4%
Logistic Regression	2-Grams	73.3%	57.0%	43.0%
MLP Neural Network	1-Grams	73.7%	53.2%	46.8%
Average			55.9%	44.1%

As observed from the results, the positive sentiment of optimism regarding the evolution and consequences of Covid-19 dominates, with a percentage exceeding 53%, which is consistently inferred across all three algorithm-encoding combinations (average 55.9%).

F. Discussion

The obtained results shed light on the sentiments expressed in Twitter users' posts and offer valuable insights into public perception regarding the Covid-19 pandemic and its variants. The analysis employed both dictionary-based and machine learning techniques to comprehensively assess sentiment dynamics.

In the realm of dictionary-based sentiment analysis, four dictionaries, namely Afinn, SenticNet, SentiWordNet, and TextBlob, were employed. Each dictionary revealed unique patterns of sentiment distribution. Notably, both TextBlob and Afinn, utilizing simplified parsing and vocabulary approaches, demonstrated similar results, where neutral emotions predominated at around 45%. Conversely, the advanced analysis methods employed by SentiWordNet and SenticNet dictionaries showcased a more nuanced sentiment landscape. SentiWordNet uncovered the coexistence of optimism and negativity, each surpassing 30%, while SenticNet revealed a clear dominance of optimism (57.61%) alongside significant negative sentiment (31.73%).

Moving to supervised machine learning sentiment analysis, three algorithms, Logistic Regression, MLP Neural Network, and Naive Bayes, were evaluated using varying N-Gram encodings. Among the encodings, the 1-Gram encoding consistently demonstrated superior efficiency for all tested algorithms, closely followed by the 2-Gram encoding. Logistic Regression with 1-Gram encoding achieved the highest accuracy of 73.9%, with the MLP Neural Network closely trailing at 73.7%. Notably, these algorithms consistently inferred a dominant positive sentiment of optimism, exceeding 53%, further reaffirming the public's relatively positive outlook on the pandemic.

When we consolidate the findings from both sentiment analysis approaches, it becomes evident that despite the evolving nature of the pandemic and the emergence of variants like Delta and Epsilon, the prevailing sentiment remains one of cautious optimism.

In conclusion, the combination of dictionary-based and machine learning techniques allowed us to gain comprehensive insights into the public sentiment surrounding the Covid-19 pandemic. The nuanced patterns uncovered by these methods provide valuable information for policymakers, researchers, and healthcare professionals, enabling them to tailor communication strategies that align with the prevailing public sentiment.

VI. Conclusions and Future Work

This study presents a systematic methodology for sentiment analysis of Twitter data related to the Covid-19 pandemic. The combination of dictionary-based and machine learning techniques provided a comprehensive understanding of sentiment dynamics and public perceptions. The findings highlight the dominance of cautious optimism among Twitter users, even amidst concerns about the pandemic's evolution.

In the future, the presented methodology can be extended to encompass sentiment analysis of other domains or specific geographic regions. Additionally, the integration of more advanced machine learning models and the exploration of deep learning techniques could yield even more nuanced insights into sentiment patterns [15], [24]. Furthermore, incorporating user demographics and contextual information might provide deeper insights into the factors shaping public sentiment.

The insights gained from sentiment analysis have applications beyond understanding public perception. They can aid in crisis management, and communication strategies during pandemics or other significant events. As social media continues to play a pivotal role in shaping public discourse, sentiment analysis remains a valuable tool for gauging sentiment trends and informing decision-making processes.

Acknowledgement

This research was funded by the European Union and Greece (Partnership Agreement for the Development Framework 2014-2020) under the Regional Operational Programme Ionian Islands 2014-2020, project title: "Indirect costs for project "TRaditional corfU Music PresErvation through digiTal innovation"", project number: 5030952.

References

[1] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau. Sentiment analysis of twitter data. In *Workshop on Language in Social Media (LSM)*, pages 30–38, 2011.

[2] W. Ahmed, J. Vidal-Alaball, J. Downing, and F. L. Seguí. Covid-19 and the 5g conspiracy theory: Social network analysis of twitter data. *Journal of Medical Internet Research*, 22(5):e19458, 2020.

[3] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *International Conference on Language Resources and Evaluation (LREC)*, 2010.

[4] R. A. Baeza-Yates and B. A. Ribeiro-Neto. *Modern Information Retrieval: The Concepts and Technology Behind Search, Second edition*. Pearson Education Ltd., Harlow, England, 2011.

[5] A. Baltas, A. Kanavos, and A. K. Tsakalidis. An apache spark implementation for sentiment analysis on twitter data. In *2nd International Workshop on Algorithmic Aspects of Cloud Computing (ALGOCLOUD)*, volume 10230, pages 15–25, 2016.

[6] E. Cambria, R. Speer, C. Havasi, and A. Hussain. Senticnet: A publicly available semantic resource for opinion mining. In *AAAI Fall Symposium*, volume FS-10-02, 2010.

[7] R. Chandra, A. Jain, and D. S. Chauhan. Deep learning via LSTM models for COVID-19 infection forecasting in india. *CoRR*, abs/2101.11881, 2021.

[8] N. F. F. da Silva, E. R. Hruschka, and E. R. H. Jr. Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, 66:170–179, 2014.

[9] G. Drakopoulos, A. Kanavos, P. Mylonas, and S. Sioutas. Discovering sentiment potential in twitter conversations with hilbert-huang spectrum. *Evolving Systems*, 12(1):3–17, 2021.

[10] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: Understanding microblogging usage and communities. In *1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, pages 56–65, 2007.

[11] E. Kafeza, A. Kanavos, C. Makris, G. Pispirigos, and P. Vikatos. T-PCCE: twitter personality based communicative communities extraction system for big data. *IEEE Transactions on Knowledge and Data Engineering*, 32(8):1625–1638, 2020.

[12] E. Kafeza, A. Kanavos, C. Makris, and P. Vikatos. T-PICE: twitter personality based influential communities extraction system. In *IEEE International Congress on Big Data*, pages 212–219, 2014.

[13] R. P. Kaila and A. V. K. Prasad. Informational flow on twitter - corona virus outbreak - topic modelling approach. *International Journal of Advanced Research in Engineering and Technology (IJARET)*, 11(3), 2020.

[14] M. Kaity and V. Balakrishnan. Sentiment lexicons and non-english languages: a survey. *Knowledge and Information Systems*, 62(12):4445–4480, 2020.

[15] A. Kanavos, F. Kounelis, L. Iliadis, and C. Makris. Deep learning models for forecasting aviation demand time series. *Neural Computing and Applications*, 33(23):16329–16343, 2021.

[16] A. Kanavos, N. Nodarakis, S. Sioutas, A. K. Tsakalidis, D. Tsolis, and G. Tzimas. Large scale implementations for twitter sentiment classification. *Algorithms*, 10(1):33, 2017.

[17] A. Kanavos, I. Perikos, I. Hatzilygeroudis, and A. K. Tsakalidis. Emotional community detection in social networks. *Computers & Electrical Engineering*, 65:449–460, 2018.

[18] A. Lyras, S. Vernikou, A. Kanavos, S. Sioutas, and P. Mylonas. Modeling credibility in social big data using LSTM neural networks. In *17th International Conference on Web Information Systems and Technologies (WEBIST)*, pages 599–606, 2021.

[19] C. Musto, G. Semeraro, and M. Polignano. A comparison of lexicon-based approaches for sentiment analysis of microblog posts. In *8th International Workshop on Information Filtering and Retrieval*, volume 1314, pages 59–68, 2014.

[20] M. Y. Ni, L. Yang, C. M. C. Leung, N. Li, X. I. Yao, Y. Wang, G. M. Leung, B. J. Cowling, and Q. Liao. Mental health, risk factors, and social media use during the covid-19 epidemic and cordon sanitaire among the community and health professionals in wuhan, china: Cross-sectional survey. *JMIR Mental Health*, 7(5):e19009, 2020.

[21] S. R. Rufai and C. Bunce. World leaders' usage of twitter in response to the covid-19 pandemic: a content analysis. *Journal of Public Health*, 42(3):510–516, 2020.

[22] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel. COVID-19 public sentiment insights and machine learning for tweets classification. *Information*, 11(6):314, 2020.

[23] A. C. Sanders, R. C. White, L. S. Severson, R. Ma, R. McQueen, H. C. A. Paulo, Y. Zhang, J. S. Erickson, and K. P. Bennett. Unmasking the conversation on masks: Natural language processing for topical sentiment analysis of covid-19 twitter discourse. *medRxiv*, 2020.

[24] A. Savvopoulos, A. Kanavos, P. Mylonas, and S. Sioutas. LSTM accelerator for convolutional object identification. *Algorithms*, 11(10):157, 2018.

[25] A. Schumacher. Textblob sentiment: Calculating polarity and subjectivity. *Planspace.org*, 7, 2015.

[26] M. Taboada, J. Brooke, M. Tofiloski, K. D. Voll, and M. Stede. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307, 2011.

[27] A. Tiwari, R. Gupta, and R. Chandra. Delhi air quality prediction using LSTM deep learning models with a focus on COVID-19 lockdown. *CoRR*, abs/2102.10551, 2021.

[28] S. Vernikou, A. Lyras, and A. Kanavos. Multiclass sentiment analysis on covid-19-related tweets using deep learning models. *Neural Computing and Applications*, 34(22):19615–19627, 2022.

[29] T. Wang, K. Lu, K. Chow, and Q. Zhu. COVID-19 sensing: Negative sentiment analysis on social media in china via BERT model. *IEEE Access*, 8:162–169, 2020.

[30] X. Wang, F. Wei, X. Liu, M. Zhou, and M. Zhang. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In *20th ACM Conference on Information and Knowledge Management (CIKM)*, pages 1031–1040, 2011.

[31] J. Xue, J. Chen, R. Hu, C. Chen, C. Zheng, Y. Su, and T. Zhu. Twitter discussions and emotions about the covid-19 pandemic: Machine learning approach. *Journal of Medical Internet Research*, 22(11):e20550, 2020.