

Natural Language Processing Implementation for Sentiment Analysis on Tweets



P. William, Anurag Shrivastava, Premanand S. Chauhan, Mudasir Raja,
Sudhir Baijnath Ojha, and Keshav Kumar

Abstract In this article, we describe our early efforts with sentiment analysis on tweets. This project is meant to extract sentiment from tweets depending on their topic matter. It utilises natural language processing methods to determine the emotion associated with a certain issue. We used three different approaches to identify emotions in our study: classification based on subjectivity, semantic association and classification based on polarity. The experiment makes advantage of emotion lexicons by establishing the grammatical relationship between them and the subject. Due to the unique structure of tweets, the proposed method outperforms current text sentiment analysis methods.

Keywords NLP · Sentiment analysis · Classification

1 Introduction

Twitter, a prominent microblogging site, enables users to publish tweets, or status updates, of up to 140 characters in length [1, 2]. These tweets often include personal opinions or sentiments about the issue being discussed. Sentiment analysis is a method for determining the user's sentiment and opinion based on their tweets. User

P. William (✉)

Department of Information Technology, Sanjivani College of Engineering, SPPU, Pune, India
e-mail: william160891@gmail.com

A. Shrivastava · P. S. Chauhan

S.D. Bansal College of Engineering, Indore, Madhya Pradesh, India

M. Raja

Department of Management Studies, University of Kashmir, Srinagar, India

S. B. Ojha

Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal (MS), India

K. Kumar

University Institute of Computing, Chandigarh University, Punjab, India

thoughts and views may be elicited in a more convenient manner than via questionnaires or surveys. The automatic extraction of sentiment from text has been the subject of a great deal of study. Using movie review domains and machine learning techniques (Naive Bayes, maximum entropy classification and support vector machine (SVM)), Pang and Lee [3–5] tested sentiment classification. Using SVM and unigram models, they were able to achieve an accuracy of up to 82.9%. However, as the performance of sentiment classification is context dependent, machine learning approaches have trouble distinguishing the emotion of text when sentiment lexicons with opposing sentiment are present. With minimum edits in graphs before sentiment classification using a machine learning technique, Pang and Lee [6] then offered the strategy of categorising texts only on their subjective content [7]. They began by determining whether or not the text included sentiment before determining whether or not the emotion was positive or negative. The accuracy was 86.4%, which was higher than in the previous trial.

Machine learning and natural language processing (NLP) methods were also featured. The polarity of sentiment lexicons may be classified using NLP to characterise the sentiment expressions connected with a certain subject. It is possible for NLP to categorise the sentiment of a text fragment rather than the whole text based merely on its subject [8]. In natural language processing, a feature extraction approach is applied. As well as collecting and correlating sentiment with certain topics, it can also extract topic-specific qualities, such as emotion, from any vocabulary that includes it. Achieving an accuracy of up to 87% for online reviews and a 9139% for general web page and news item ratings, it beats machine learning methods. In order to produce a better result, this technique focused on the overall text and deleted particular problematic conditions such as confused words or sentences that lack emotion, for instance [9, 10].

For sentiment analysis of text, previous machine learning and natural language processing research may not be relevant to tweets because of their structural peculiarities. Twitter sentiment analysis is unique from previous textual research in three ways: The size of the object. The maximum character count for a tweet is 140. Tweets are often only 14 words in length, and sentences are typically 78 characters long, according to research by Go and colleagues [11]. Sentiment analysis in tweets and text is separate since tweets are shorter in duration while text sentiment analysis focuses on long review articles. Easily accessible data. When comparing tweets with regular text, the amount of information is different. Pang and Lee [12–15] employed a 2053-word corpus for training and testing and categorised feelings using machine learning algorithms. But, for their work on Twitter sentiment analysis, Go et al. [16–18] gathered up to 15,000 tweets of sentiment. We can now collect tens of thousands or even millions of tweets for training thanks to the Twitter API. The sentence's will give different verticals to various organisations. Acronyms, abbreviations and long sentences abound in tweets, resulting in a disjointed language. In addition to text, emojis, a URL, a photo, hashtags, punctuation and more may be included in the message. Since these components aren't actual words that can be found in a dictionary or read and comprehended by a computer, they detract from the analysis

process's accuracy. Since robots are unable to understand informal languages, some method must be developed [19, 20].

To better understand how people communicate, researchers combined grammatical analysis with word frequency analysis. Grammatical analysis looked at the structure of the text and established a relationship between the emotion lexicons and the subject in order to tie them to the topic [21, 22]. There has been a significant leap forward in sentiment analysis for short colloquial texts because older techniques were unable to identify sentiment accurately. In spite of the fact that it didn't need any supervised teaching, this method boosted previous job accuracy by 40%. The goal of this project is to provide a mechanism for analysing the sentiment of tweets in relation to a certain topic. Many pre-processing methods were used to reduce noise from tweets and show them in a more formal language. It is possible to determine the sentiment of tweets by analysing the content of the tweets and using natural language processing to detect and categorise the sentiment of the tweets. Tweets will be labelled as either positive or negative or as neutral as possible. Here is the remainder of the essay: Sect. 2 provides an overview of the proposed system's architecture and the test data set. When comparing a suggested system's performance to that of current tools, Sect. 3 analyses the experimental data, and Sect. 4 summarises the work's conclusion.

2 Overview of Framework

This section explains the architecture of the proposed system. Tweets were retrieved from a Twitter database for the experiment. Each tweet was meticulously categorised as either positive, negative or neutral based on how it was received. As a way to evaluate the proposed system's accuracy and precision, this set of tweets was employed. Pre-processing of the dataset was necessary before the suggested approach could be used to analyse the tweets. In order to ensure that robots can read and comprehend tweets, pre-processing is required. Sentiment classification may be used to determine the emotional tone of tweets after pre-processing. Subjectivity, semantic association and polarity are the three components of sentiment categorization. After determining whether tweets were subjective or objective, semantic association was utilised to find sentiment lexicons that were connected with the subject matter. Positive, negative or neutral emotion lexical categorization predicted whether tweets were positive, negative or neutral [23–25].

Data gathering more than 1500 tweets were manually tagged on Twitter and then extracted for this study. Tweets with the hashtag "Unifi" allude to a Malaysian telecommunications company. Furthermore, it is used to classify people's emotions about it [26]. There are 345 tweets that are positive, 641 that are negative and 531 that are neutral. The proposed method analyses tweets in order to predict future emotion. Alchemy API1 and Weka2 were used to analyse 1513 tweets for benchmarking purposes. Alchemy API uses natural language processing to analyse sentiment, while Weka uses machine learning techniques to mine data. As far as machine

learning methods go, we settled on Naive Bayes, decision tree (J48) and support vector machines. Raw and pre-processed tweets were both imported into Weka, and the results were compared to see how much of a difference pre-processing made. In Weka, features are extracted using an algorithm. Cross-validation was used to train and assess the data by selecting the top 100 terms and doing a tenfold cross-validation. Results from Alchemy API and Weka were combined with the manually labelled tweets in order to calculate accuracy, precision, recall and F-measure B. Assembled proposal as seen in this graphic, the suggested system's steps begin with pre-processing and culminate with sentiment classification [27–30].

Pre-processing is covered in Sect. 1, while the emotion classification approach is explained in detail in this section. Because most tweets are unstructured text, pre-processing is employed to organise and present them. It also helps machines better understand the content of the tweets [31–33]. Replace special symbols, extend abbreviations and acronyms and capitalise topics by removing URL and #hashtags [34].

For the sake of brevity, URLs and image links are not included in the text. There are no hashtags in this text to avoid confusion, as a hashtag may not be directly linked to the subject. By using words instead of specific symbols, the text processing process is made simpler. For example, '>' is replaced with 'greater', and '&' is replaced with 'and'. Word-based sentiment analysis beat emoticon-based sentiment analysis in research on automatic sentiment analysis of Twitter messages.

Because of this, emoticons in tweets are disabled. In order to make unstructured tweets more readable, it is common practise to abbreviate long words like 'good' to 'good.' Abbreviations, acronyms, and contractions have all had their letters and numbers increased in size. The phrase "I'm not going to work 2mr," for example, may be expanded to mean "I'm not going to work the next day." To make it easier for the machine to read and understand your text, you should utilise topic capitalization. Sentiment classification will be used to the processed tweets in order to forecast their sentiment. Sentiment Classification: The sentiment classification process is shown in Figs. 1 and 2.

- (a) **Subjectivity Classification:** It is possible to categorise tweets into subjective or objective categories using subjective categorization. The programme analyses each tweet word by word to determine whether or not it contains any emotive language. The message will be classified as subjective if the phrase used in the tweet evokes either good or negative feeling. Otherwise, it will be objective, which is neutral. Alternatively, it will be subjective. A good example of this would be, "Come acquire an internet package" or "Come get a new internet plan." The first tweet does not include a phrase that indicates an emotional rating. Objectivity and neutrality are assigned to it. The adjective "new" is used in the second tweet to convey positive feelings. Substantive association will be performed on the tweet before classifying it as subjective.
- (b) **Semantic Association:** In semantic association, grammatical linkages between the topic and sentiment lexicons are used to identify sentiment lexicons that are relevant to the subject. There are fewer rules to follow while composing a tweet

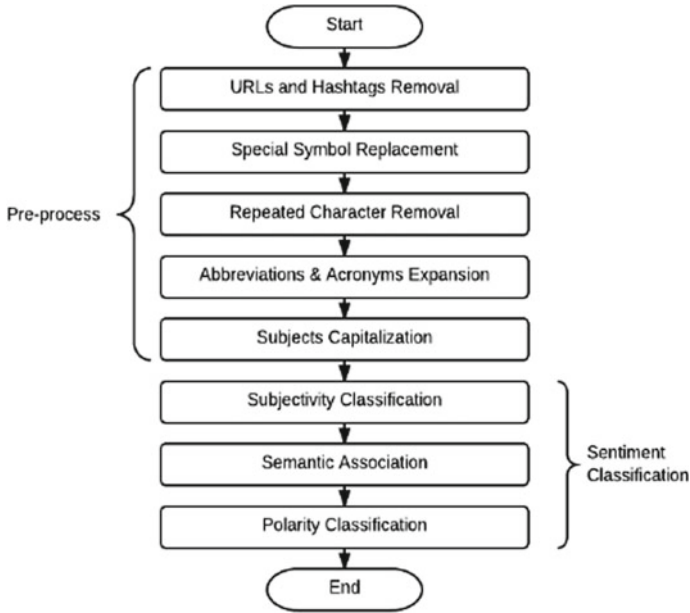


Fig. 1 Sentiment classification flowchart

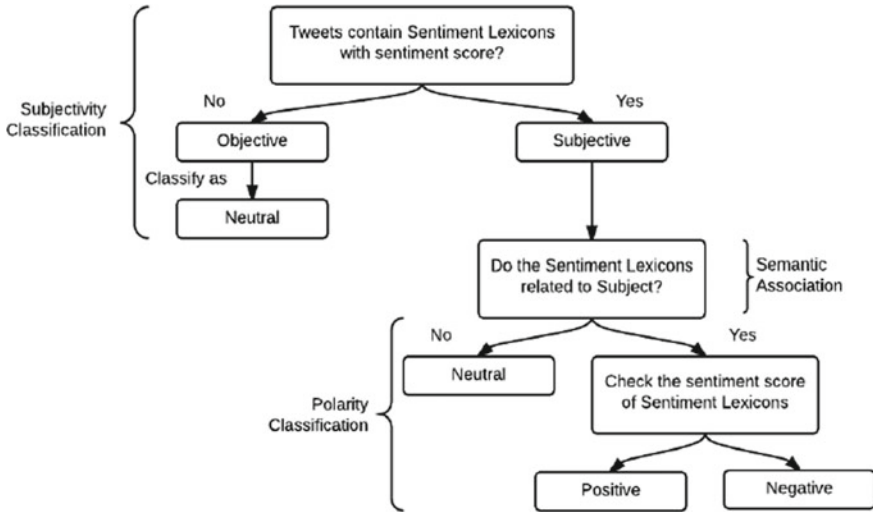


Fig. 2 Processing of sentiment classification

since they are shorter and more to-the-point. Adjectives and verbs are two types of sentiment lexicons that are often used in conjunction with a subject. Tweets [12] may provide first-person perspectives and side-by-side comparisons.

Using prepositions and conjunctions, emotion lexicons describe one or more topics in a direct opinion. In contrast, there are at least two subjects in an opinion, but the subjects are linked to the same emotional lexicons without the existence of a conjunction. They are connected. You may see “I love Unifi” as a simple statement at Alg. 1. If ‘I’ is the nominal subject and ‘Unifi’ is the direct object of ‘love’, then the illustration illustrates the relationship between the two. It is a straightforward statement, and this is how the majority of tweets are phrased.

The bulk of grammatical connections show that verbs and adjectives are linked to the subject, as described earlier. ‘Unifi’ is the direct object of the attribute ‘love’ in this instance; thus, we must validate it. To summarise, the POS tag shows that “I,” “love” and “Unifi” are all nouns. “Love” will be classified as either positive or negative since it is a verb that connects with the subject of the sentence.

It is important to look for grammatical patterns that are related to the subject matter:

Adjective used to describe a subject (good Unifi finally upgrade the service).

Adjective or verb that is attached to a noun or pronoun (happy when Unifi is recovered).

Adverbs that describe the subject matter (Unifi speed is fine).

Between it and the subject, there is an adjective with a preposition (fast like Unifi).

Adjective with a preposition between it and the subject that is superlative in nature (let us face it Unifi is not the best but it is better than M). Verbs, adjectives or nouns used in connection with the subject matter (50% of my draughts are about Unifi to be honest). Adjectives and verbs that describe the subject matter (Unifi forever no lag). To invert an emotion, use a negate word in the adjective or verb (I do not want to uninstall Unifi). However, the sentence structure and grammatical connections of comparison opinion are unique. Algorithm 2 provides a comparative assessment of the concepts in the text. It is better than M.

In this case, the nominal is used instead. Prepositions are used between the subject of ‘better’ and another word called ‘M’, and ‘Unifi’ is the object of comparison. For example, the adjective “Unifi is better” may be seen as a grammatical pattern, as can the preposition “better than” in the superlative adjective “better than M.” In this section, we will discuss polarity classifications. Polarity classification is used to classify tweets based on their subjective content. Sentiment lexicons related with certain topics are used to classify tweet sentiment.

To provide an example, when someone says, “I love Unifi,” they are using the emotion lexicon. SentiWordNet reports that “love” gets a score of 0.625%. Because of this, we may conclude that “Unifi” is feeling happy and hence label the tweet as such [13].

Comparative opinion relies on the subject’s viewpoint. Even though the tweet ‘Unifi is better than M’ includes two subjects—‘Unifi’ and ‘M’, the adjectival term “better” appears. When a comparative adjective precedes a topic, it expresses a

different sentiment than the previous subject. Unfi will be scored positive in this circumstance since ‘better’ has a positive score of 0.825%, whereas M will be rated negative.

Algorithm of Dependencies Type and POS Type of Direct Opinion

Sentence *I love Unfi*

Pos Tagging

I/PRP
love/VBP
Unfi/NNP

Parse

(ROOT
(S
(NP (PRP I))
(VP (VBP love)
(NP (NNP Unfi))))))

Typed dependencies

nsubj(love-2, I-1)
root(ROOT-0, love-2)
dobj(love-2, Unfi-3)

Algorithm of Dependencies Type and POS Type of Comparison Opinion

Sentence *Unfi is better than M*

Pos Tagging

Unfi/NNP
is/VBZ
better/JJR
than/IN
M/NNP

Parse

(ROOT
(S
(NP (NNP Unfi))
(VP (VBZ is)
(ADJP
(ADJP (JJR better))
(PP (IN than)
(NP (NNP M))))))

Typed dependencies, collapsed

nsubj(better-3, Unfi-1)
cop(better-3, is-2)
root(ROOT-0, better-3)
prep_than(better-3, M-5)

3 Result and Discussions

Findings have been summarised in an incoherent matrix of confusion. Both predicted and actual outcomes are recorded. Confusion matrix is of size $\ell \times \ell$, where ℓ is the number of different label values [6]. Positive, negative and neutral labels are employed in this research. We may calculate the accuracy, recall and F-measure scores by comparing the predicted results to the actual ones.

Table 1 summarises the performance of the proposed system and the Alchemy API. With an accuracy of 59.85%, a precision of 53.65% and F-measurement of 0.48, the proposed system outperforms the Alchemy API. Alchemy API has an F-measure of 0.43 and an accuracy, precision and precision of 58.87%. Since tweets have a different structure than ordinary text, Alchemy API may not be able to accurately analyse their sentiment. Algorithm performance is summarised in Table 2. Pre-processed tweets provide 64.95% accuracy, 66.54% precision and 0.57 in the F-measure, whereas raw tweets yield 58.67% accuracy, 60.44% precision and 0.48 in the F-measure. Even when using pre-processed tweets, it was able to outperform Naive Bayes and decision tree classifiers and come out on top in the end. According to Weka’s research, NLP-based pre-processing greatly improves performance when compared to those that use raw tweets as a corpus. ‘Classifier accuracy, precision, and F-measure’ dropped on average by 2.09%, 5.23% and 0.10 in pre-processed tweets when trained and assessed. Alchemy API and SVM are contrasted in Table 3 based on their respective performance. The Alchemy API is beaten in general but not SVM. SVMs perform better when they are trained and tested on tweets that have been pre-processed. Consequently, the proposed system has to be improved to reach a greater degree of performance.

Table 1 Result from proposed system and Alchemy API

	Proposed system	Alchemy API
Accuracy	59.85%	58.87%
Precision	53.65%	40.82%
F-Measure	0.48	0.43

Table 2 Result from Weka using machine learning algorithm

	Naïve Bayes		Decision tree		Support vector machine	
	Raw tweets	Pre-processed tweets	Raw tweets	Pre-processed tweets	Raw tweets	Pre-processed tweets
Accuracy	55.04%	60.58%	49.70%	57.60%	58.67%	64.95%
Precision	47.47%	53.59%	44.05%	53.65%	60.44%	66.54%
F-Measure	0.44	0.55	0.35	0.43	0.48	0.57

Table 3 Summarised result for 3 different sentiment analysis tools

	Proposed system	Alchemy API	SVM (Pre-Processed Tweets)
Accuracy	59.85%	58.87%	64.95%
Precision	53.65%	40.82%	66.54%
F-Measure	0.48	0.43	0.57

4 Conclusion and Future Work

There is a wealth of study on extracting emotion from tweets, owing to Twitter's popularity as a social media site. We provide early findings for our proposed system, which incorporates natural language processing techniques to extract topic from tweets and classifies tweets' polarity using sentiment lexicons linked with subject.

SVM outperforms Alchemy API in the tests, while the proposed system exceeds Alchemy API. Research in this area will focus on ways to make sentiment analysis more accurate. The use of slang and misspelt words makes it difficult to derive emotion lexicons from tweets that haven't been formalised beforehand. Due to the need for additional training data, pre-processing needs turning tweets into formal phrases, which is still inefficient.

References

1. William P, Badholia A (2021) Analysis of personality traits from text based answers using HEXACO model. In: 2021 International conference on innovative computing, intelligent communication and smart electrical systems (ICSES), pp 1–10. <https://doi.org/10.1109/ICS ES52305.2021.9633794>
2. William P, Badholia A (2021) Assessment of personality from interview answers using machine learning approach. *Int J Adv Sci Technol* 29(08):6301–6312
3. William P, Badholia A (2020) Evaluating efficacy of classification algorithms on personality prediction dataset. *Elementary Educ Online* 19(4):3400–3413. <https://doi.org/10.17051/ilkonline.2020.04.764728>
4. William P, Badholia A, A review on prediction of personality traits considering interview answers with personality models. *Int J Res Appl Sci Eng Technol (IJRASET)* 9(V):1611–1616. ISSN: 2321-9653
5. William P, Patil VS (2016) Architectural challenges of cloud computing and its security issues with solutions. *Int J Sci Res Develop* 4(8):265–268
6. William P, Kumar P, Chhabra GS, Vengatesan K (2021) Task allocation in distributed agile software development using machine learning approach. In: 2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON), pp 168–172. <https://doi.org/10.1109/CENTCON52345.2021.9688114>
7. William P, Badholia A, Verma V, Sharma A, Verma A (2022) Analysis of data aggregation and clustering protocol in wireless sensor networks using machine learning. In: Suma V, Fernando X, Du KL, Wang H (eds) *Evolutionary computing and mobile sustainable networks*. Lecture Notes on Data Engineering and Communications Technologies, vol 116. Springer, Singapore. https://doi.org/10.1007/978-981-16-9605-3_65

8. Bibave R, Thokal P, Hajare R, Deulkar A, William P, Chandan AT (2022) A comparative analysis of single phase to three phase power converter for input current THD reduction. In: 2022 International conference on electronics and renewable systems (ICEARS), pp 325–330. <https://doi.org/10.1109/ICEARS53579.2022.9752161>
9. Bornare AB, Naikwadi SB, Pardeshi DB, William P (2022) Preventive measures to secure arc fault using active and passive protection. In: 2022 International conference on electronics and renewable systems (ICEARS), pp 934–938. <https://doi.org/10.1109/ICEARS53579.2022.9751968>
10. Pagare KP, Ingale RW, Pardeshi DB, William P (2022) Simulation and performance analysis of arc guard systems. In: 2022 International conference on electronics and renewable systems (ICEARS), pp 205–211. <https://doi.org/10.1109/ICEARS53579.2022.9751924>
11. Matharu HS, Girase V, Pardeshi DB, William P (2022) Design and deployment of hybrid electric vehicle. In: 2022 International conference on electronics and renewable systems (ICEARS), pp 331–334. <https://doi.org/10.1109/ICEARS53579.2022.9752094>
12. William P, Choubey A, Chhabra GS, Bhattacharya R, Vengatesan K, Choubey S (2022) Assessment of hybrid cryptographic algorithm for secure sharing of textual and pictorial content. In: 2022 International conference on electronics and renewable systems (ICEARS), pp 918–922. <https://doi.org/10.1109/ICEARS53579.2022.9751932>
13. William P, Choubey S, Ramkumar M, Verma A, Vengatesan K, Choubey A (2022) Implementation of 5G network architecture with interoperability in heterogeneous wireless environment using radio spectrum. In: 2022 International conference on electronics and renewable systems (ICEARS), pp 786–791. <https://doi.org/10.1109/ICEARS53579.2022.9752267>
14. Pawar AB, Gawali P, Gite M, Jawale MA, William P (2022) Challenges for hate speech recognition system: approach based on solution. In: 2022 International conference on sustainable computing and data communication systems (ICSCDS), pp 699–704. <https://doi.org/10.1109/ICSCDS53736.2022.9760739>
15. William P, Jadhav D, Cholke P, Jawale MA, Pawar AB (2022) Framework for product anti-counterfeiting using blockchain technology. In: 2022 International conference on sustainable computing and data communication systems (ICSCDS), pp 1254–1258. <https://doi.org/10.1109/ICSCDS53736.2022.9760916>
16. William P, Gade R, Chaudhari R, Pawar AB, Jawale MA (2022) Machine learning based automatic hate speech recognition system. In: 2022 International conference on sustainable computing and data communication systems (ICSCDS), pp 315–318. <https://doi.org/10.1109/ICSCDS53736.2022.9760959>
17. William P, Badholia A, Patel B, Nigam M (2022) Hybrid machine learning technique for personality classification from online text using HEXACO model. In: 2022 International conference on sustainable computing and data communication systems (ICSCDS), pp 253–259. <https://doi.org/10.1109/ICSCDS53736.2022.9760970>
18. Pawar AB, V. Khemnar, R. Londhe, P. William and M. A. Jawale, “Discriminant Analysis of Student’s Online Learning Satisfaction during COVID’ 19,” 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022, pp. 260–263. <https://doi.org/10.1109/ICSCDS53736.2022.9760895>
19. Yuvaraj S, Badholia A, William P, Vengatesan K, Bibave R (2022) Speech recognition based robotic arm writing. In: Goyal V, Gupta M, Mirjalili S, Trivedi A (eds) Proceedings of international conference on communication and artificial intelligence. Lecture Notes in Networks and Systems, vol 435. Springer, Singapore. https://doi.org/10.1007/978-981-19-0976-4_3
20. Gondkar SS, Pardeshi DB, William P (2022) Innovative system for water level management using IoT to prevent water wastage. In: 2022 International conference on applied artificial intelligence and computing (ICAAIC), pp 1555–1558. <https://doi.org/10.1109/ICAAIC53929.2022.9792746>
21. Wakchaure A, Kanawade P, Jawale MA, William P, Pawar AB (2022) Face mask detection in realtime environment using machine learning based google cloud. In: 2022 International conference on applied artificial intelligence and computing (ICAAIC), pp 557–561. <https://doi.org/10.1109/ICAAIC53929.2022.9793201>

22. Kolpe R, Ghogare S, Jawale MA, William P, Pawar AB (2022) Identification of face mask and social distancing using YOLO algorithm based on machine learning approach. In: 2022 6th International conference on intelligent computing and control systems (ICICCS), pp 1399–1403. <https://doi.org/10.1109/ICICCS53718.2022.9788241>
23. Batt AA, Ahmad Bhat R, Pardeshi DB, William P, Gondkar SS, Singh Matharu H (2022) Design and optimization of solar using MPPT algorithm in electric vehicle. In: 2022 6th International conference on intelligent computing and control systems (ICICCS), pp 226–230. <https://doi.org/10.1109/ICICCS53718.2022.9787988>
24. Najgad YB, Namedev Munde S, Chobe PS, Pardeshi DB, William P (2022) Advancement of hybrid energy storage system with PWM technique for electric vehicles. In: 2022 6th International conference on intelligent computing and control systems (ICICCS), pp 238–242. <https://doi.org/10.1109/ICICCS53718.2022.9788135>
25. Ghoderao RB, Raosaheb Balwe S, Chobe PS, Pardeshi DB, William P (2022) Smart charging station for electric vehicle with different topologies. In: 2022 6th International conference on intelligent computing and control systems (ICICCS), pp 243–246. <https://doi.org/10.1109/ICICCS53718.2022.9788143>
26. Gondkar SS, William P, Pardeshi DB (2022) Design of a novel IoT framework for home automation using google assistant. In: 2022 6th International conference on intelligent computing and control systems (ICICCS), pp 451–454. <https://doi.org/10.1109/ICICCS53718.2022.9788284>
27. William P et al (2022) Darknet traffic analysis and network management for malicious intent detection by neural network frameworks. In: Rawat et al (eds) Using computational intelligence for the dark web and illicit behavior detection. IGI Global, pp 1–19. <https://doi.org/10.4018/978-1-6684-6444-1.ch001>
28. William P et al (2022) Systematic approach for detection and assessment of dark web threat evolution. In: Rawat et al (eds) Using computational intelligence for the dark web and illicit behavior detection. IGI Global, pp 230–256. <https://doi.org/10.4018/978-1-6684-6444-1.ch013>
29. Blenn N, Charalampidou K, Doerr C (2012) Context-sensitive sentiment classification of short colloquial text. In: Networking 2012, pp 97–108. Springer, Berlin, Heidelberg
30. Davenport SW, Bergman SM, Bergman J Z, Fearington ME (2014) Twitter versus Facebook: exploring the role of narcissism in the motives and usage of different social media platforms. *Compute Hum Behav* 32:212–220
31. Esuli A, Sebastiani F (2006) Sentiwordnet: a publicly available lexical resource for opinion mining. *Proc. LREC* 6:417–422
32. Go A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, pp 1–12
33. Go A, Huang L, Bhayani R (2009) Twitter sentiment analysis. *Entropy* 17
34. Lima C, de Castro LN (2012) Automatic sentiment analysis of Twitter messages. In: Computational aspects of social networks (CASoN), 2012 Fourth International Conference. IEEE, pp 52–57