Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 3, 145-155
2024
Publisher: Learning Gate
DOI: 10.55214/25768484.v8i3.847

Sentiment analysis of the awareness of environmental sustainability

Kunyanuth Kularbphettong[1*], Pattarapan Roonrakwit[2], Chongrag Boonseng[3]
[1]Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, Thailand; kunyanuth.ku@ssru.ac.th (K.K.).
[2]Faculty of Information and Communication Technology, Silpakorn University, Bangkok, Thailand; ajpui20@hotmail.com (P.R.).
[3]School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand; chongrag.bo@kmitl.ac.th (C.B.).

Abstract: This study examines the sentiment analysis of awareness of environmental sustainability. Environmental sustainability is the responsible management and utilization of Earth's natural resources to meet the needs of the present generation and ensure that future generations will access those resources. The awareness of environmental sustainability has been growing globally as people, businesses, and governments recognize the importance of preserving the planet for current and future generations. Sentiment analysis of environmental sustainability involves evaluating opinions, attitudes, and emotions expressed in texts related to environmental sustainability, and analyzing sentiment can provide insights into public perception, awareness, and engagement with environmental issues. This exploratory study's primary goal is to conduct social media opinion mining in the context of Thai people's environmental sustainability. The paper presented how to build a model of sentiment analysis with linguistic analysis, including data preprocessing steps, feature extraction, and model constructions. The techniques used in this research include Logistic Regression, Random Forests, Support Vector Machine, Word Segmentation and Bag of Words. The result shows that the model is able to categorize sentiment analysis opinions in the sustainability context primarily in positive terms. The positive sentiments suggest a sustained, long-term shift in awareness, or they might be influenced by specific events or trends. However, positive sentiment analysis results are expressed towards environmental sustainability initiatives, such as renewable energy projects, waste reduction efforts, or conservation programs. Moreover, public awareness plays a crucial role in influencing individual behavior, corporate practices, and government policies towards a more sustainable and environmentally conscious future.
Keywords: *Awareness, Environmental sustainability, Logistic regression, Machine learning, Random forests, Sentiment analysis, Support vector machine.*

1. Introduction

Nowadays, the overexploitation of natural resources is one of the significant problems that occur all over the world, and degradation of the natural resources is driving the three crucial impacts, including climate change, biodiversity loss, and pollution [1, 2]. Climate change refers to the long-term alterations in worldwide or regional weather patterns, encompassing variations in temperature, precipitation, and other measures. The consequences of climate change are wide-ranging and include rising sea levels, more frequent and intense heatwaves, droughts, floods, and storms, along with disruptions to ecosystems and the extinction of species. Climate change also has significant economic and social impacts, affecting agriculture, energy, health, and migration patterns, among other areas. Biodiversity loss refers to the decline in the variety and abundance of plant and animal species, and the loss of biodiversity has far-reaching consequences for human well-being, as well as the health and functioning of ecosystems and economic development. Pollution is the presence of harmful pollutants in

History: Received: 2 January 2024; Revised: 20 February 2024; Accepted: 6 March 2024; Published: 29 March 2024
* Correspondence: kunyanuth.ku@ssru.ac.th

the natural environment, such as air, water, or soil. Pollution is caused by various human activities, including industrial and agricultural practices, transportation, and waste disposal, and these different types of pollution can have harmful effects on human health, wildlife, and the environment.

Thailand is a country where environmental deterioration is on the rise in many areas. These include deforestation, desertification, water scarcity, climate change, falling wildlife populations, and pollution of the air and water [3, 4]. According to Worldometer [5] Thailand was ranked 22nd in terms of national CO_2 emissions. Also, Thailand falls 11 ranks to 42nd in this year's CCPI (the Climate Change Performance Index) from a medium to a low performer in GHGEmissions (Greenhouse gasesEmissions), Renewable Energy, and Climate Policy [6]. Overall, Thailand faces significant environmental sustainability challenges, but the government has implemented initiatives to address these issues and promote sustainability. However, much more needs to be done to ensure a sustainable future for the country and its people. Environmental sustainability is the responsible use and conservation of natural resources in order to maintain a healthy environment for both present and future generations. The important aspects of environmental sustainability include conserving natural resources, reducing pollution, minimizing waste and promoting biodiversity. Therefore, awareness of environmental sustainability is crucial because it helps people understand the impact of their actions on the environment and motivates them to make changes to reduce their environmental footprint. There are many approaches to increasing awareness of environmental sustainability, like educating people about environmental issues, encouraging sustainable practices, and implementing policies and regulations by governments.

The media plays a crucial role in shaping public opinion and increasing awareness. News outlets can cover environmental issues and provide information on how individuals can take action. At present, social media platforms are a rich source of information about people's attitudes. To analyze and monitor public opinion through social media, sentiment analysis is a contextual mining method used in natural language processing to ascertain the tone or viewpoint of a given text, like a customer review or a social network post [7]. There are many ways sentiment analysis can be used to understand attitudes towards environmental sustainability: Brand perception: Sentiment analysis can help businesses understand how their brand is perceived in terms of environmental sustainability by analyzing customer reviews and social media posts. Public opinion: Sentiment analysis can also be used to understand public opinion on environmental issues, such as climate change or plastic pollution, by analyzing social media posts and news articles. Campaign evaluation: Sentiment analysis can be used to evaluate the effectiveness of environmental sustainability campaigns by analyzing social media posts and other data and Industry analysis: Sentiment analysis can also be used to understand attitudes towards sustainability within specific industries by analyzing language data from industry-specific sources. With the recent advances in deep learning, sentiment analysis algorithms have improved considerably. Hence, sentiment analysis is a valuable tool for understanding public attitudes towards environmental sustainability and identifying areas for improvement.

2. Literature Review

This section provides a concise overview of the relevant studies that have utilized machine learning techniques for sentiment similarity. In recent years, there has been significant research on opinion mining and sentiment analysis to ascertain and summarize sentiment polarity.

Awan, et al. [8] proposed several machine learning models, including the random forest classifier, logistic regression, and the term frequency-inverse document frequency (TF-IDF) for text vectorization, which have demonstrated the ability to effectively identify fraudulent news. The accuracy of these models makes them a valuable tool for obtaining results that are grounded in reality and can be applied to diverse, unstructured data for various sentiment analysis applications. The combination of TF-IDF and the cosine similarity technique was employed to gauge sentiment similarity, achieving an accuracy of 81.2% [8]. TF-IDF was employed to extract features from comment reviews, followed by the application of an ensemble method that incorporated Support Vector Machines (SVMs) to classify the sentiment of the reviews [9-12].

According to Singh and Tripathi [13] a comparative analysis of the Support Vector Machine (SVM), Random Forest, and Decision Tree algorithms was conducted, wherein their performance was evaluated based on accuracy, recall, precision, and F1 measures. It was determined that the Decision Tree algorithm exhibited the highest level of accuracy. The ensemble classifier is trained on features like lexical to determine the polarity of each tweet, Clark and Wicentwoski [14]. Bahrawi [15] collected emotions or 'feelings' data from Twitter and applied the Random Forest algorithm, which resulted in a high level of accuracy. In addition, the data from an online shopping website was analyzed using the Random Forest and Support Vector Machine algorithms [16]. Another method proposed for sentiment classification is the utilization of cosine similarity, to classify the sentiment expressed by a user's comment [17]. Furthermore, there exist a significant number of related studies [18-21] that primarily focus on mining customer reviews and determining the polarity of the opinions expressed in them.

3. Experimental Setup

3.1. Data Sets and Data Preprocessing

There are a multitude of sources and approaches by which to acquire data for the purpose of analyzing environmental sustainability. In order to obtain data for a comprehensive examination, it is possible to glean valuable insights regarding public attitudes and behaviors through the utilization of social media platforms such as Twitter, YouTube, Facebook, and Instagram.

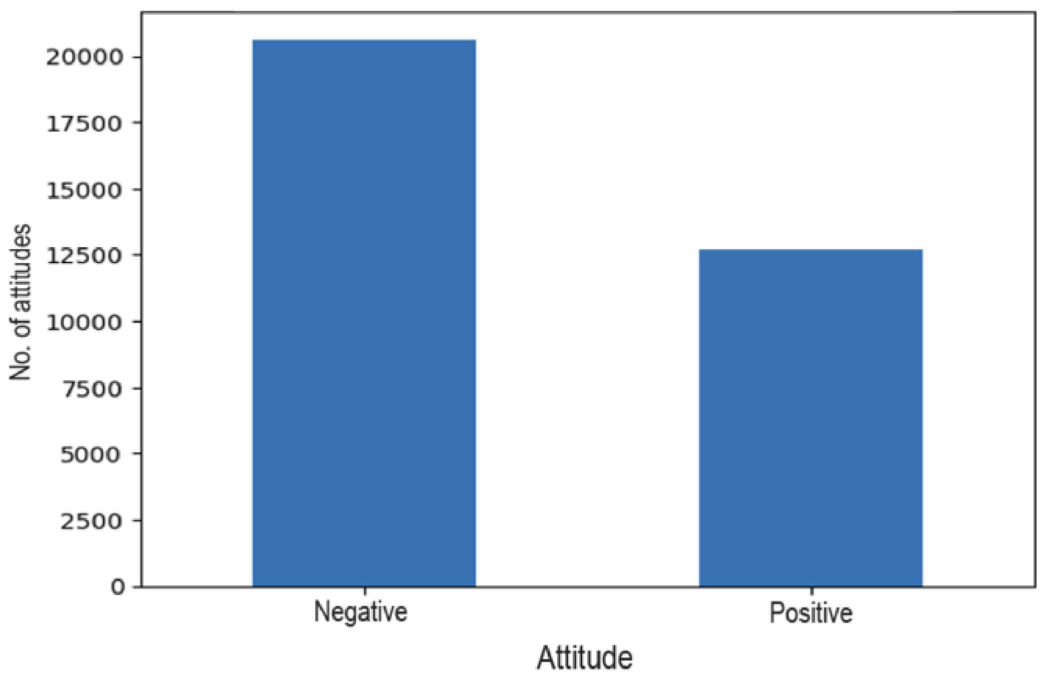

Figure 1.
The number of positive and negative attitudes.

Data collected through web scraping expresses opinions about environmental sustainability. Twitter's rebranded identity, named "X," was conceived by Elon Musk, who assumed ownership of Twitter in 2022.X's APIs, or Twitter API (Twitter Application Programming Interface), and in order to retrieve the tweets, the Tweepy API [22] is used, and data is recorded in the database, including Twitter, hashtags, created tweet times, tweet texts, and retweet counts. As seen in Figure 1, the data

DOI: 10.55214/25768484.v8i3.847

exploration phase is utilized to examine and analyze the data in order to find trends and insights from both positive and negative perspectives.

Text mining requires data preparation, which involves converting unstructured textual data into a clean, organized format suitable for analysis [23-25]. However, the Thai language differ from English in that there are no markings and symbols to indicate the scope of each word or sentence.

The following are related techniques used in data preprocessing for Thai text mining:

1. Word segmentation: Thai does not use spaces between words, so the first step in preprocessing Thai text data is to segment the text into individual words. This is typically done using a word segmentation tool.

2. Symbol and Emoji Removal: symbols, like '@#$& []', are removed from the reviews, and emoji and emoticons, often expressed post feelings, are eliminated from the data set.

3. Stop word removal: Similar to other languages, there are several frequent terms in Thai, including pronouns and particles that have little meaning. Eliminating these stop words can aid in decreasing the dataset's size and enhancing the analytical precision [26].

4. Normalization: There are numerous words in Thai with similar meanings but different spellings. To convert the text into a standard format, the Pythainlp package [27] which was created using the Python® language, was utilized by converting these terms to their standard form. Normalizing approach can help decrease the dataset's size and increase the analysis accuracy.

5. Removing HTML (Hypertext Markup Language) tags and URLs(Uniform Resource Locator): Removing any HTML tags or URLs that may be present in the text.

6. Handling misspellings and abbreviations: correcting common misspellings and expanding abbreviations to their full form. Data preprocessing in text mining is an essential step that helps to ensure accurate and meaningful analysis of textual data.

3.2. Data Processing

Analyzing the awareness of environmental sustainability in sentiment can be done by following these general steps:

1. Collect data: Collect data on public awareness and attitudes towards environmental sustainability using the methods mentioned in the section above.

2. Preprocess data: preprocess the data by cleaning it and transforming it into a format that can be analyzed. This may include removing irrelevant data, standardizing the format of the data, and converting it to a format that can be analyzed.

3. Sentiment analysis: Use a sentiment analysis tool to analyze the sentiment of the data. This involves analyzing the text data and categorizing it as positive or negative using machine learning. To determine whether a sentiment score is positive or negative, a threshold value is set to separate positive from negative sentiment. For example, a threshold of 1 is considered positive, and any score below -1 is considered negative. The proposed steps of a sentiment analysis model using machine learning are as follows:

1. Split the data into training and testing sets: This involves dividing the labeled dataset into two sets: one for training the model and one for evaluating the performance of the model.

2. Train a machine learning model: This involves selecting an appropriate algorithm and training it on the training data.

2.1 TF/IDF Vectorization: The next step is to convert data into a numerical vector format using the TF/IDF (Term Frequency-Inverse Document Frequency) vectorization technique. The approach assigns a weight to each term in the document based on its frequency and how often it appears in other documents.

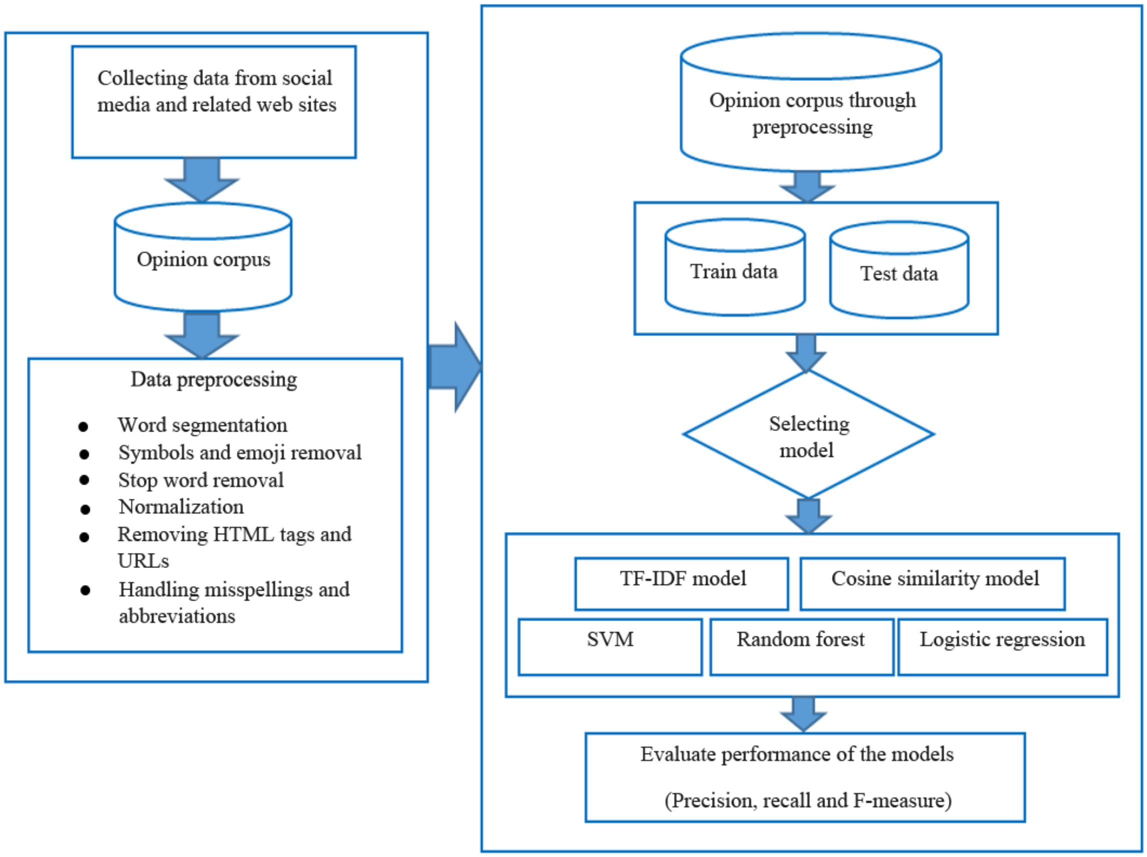

Figure 2.
The system architecture of proposed approach.

2.2 Cosine Similarity: After vectorizing the text data, cosine similarity was used to measure the similarity between the documents. Cosine similarity calculates the cosine of the angle between two vectors.

2.3 Evaluate the performance of the model: This involves testing the trained model on the testing data and evaluating its performance using metrics such as accuracy, precision, recall, and F1 score. Figure 2 shows the system architecture representing the steps used in our approach.

3.3. Methodologies

Sentiment analysis can be used to analyze attitudes towards environmental sustainability, including both positive and negative sentiment, and this section explains the methodologies used in this research. SVM, Random Forest, and Logistic Regression are all popular machine learning algorithms used for classification tasks.

3.3.1. TF-IDF

The concept of TF-IDF (Term Frequency-Inverse Document Frequency) entails the utilization of a numerical metric to assess the significance of a word in a document or a collection of documents. The term frequency (TF) gauges the frequency at which a word appears in a specific document, while the inverse document frequency (IDF) gauges the rarity of a word across all documents in the corpus. The underlying notion behind IDF is that words that are frequently employed in numerous documents possess less significance in comparison to words that are rarely used and only found in a select few documents. The formula utilized for computing the TF-IDF of a term in a document is:

$$TF - IDF = TF * IDF \, , \tag{1}$$

Where TF is the number of times the term appears in the document and IDF is calculated as follows:

$$IDF = log \, (\frac{N}{n}), \tag{2}$$

Where N is the total number of documents in the corpus, and n is the number of documents in which the term appears.

TF-IDF is commonly used in information retrieval and text mining applications to find relevant documents based on a user's search query [28]. The higher the TF-IDF score of a term in a document, the more relevant that document is likely to be to the search query.

3.3.2. Cosine Similarity

Cosine similarity serves as an indicator of likeness between two non-zero vectors within a multi-dimensional space. It gauges the cosine of the angle formed between the two vectors, thereby providing insight into the extent to which their orientations align. In the realm of natural language processing, cosine similarity frequently serves as a means to measure similarity between two documents or texts. The text is represented as a vector of word frequencies, and subsequently, the cosine similarity between two text vectors is calculated to ascertain the level of similarity between the two texts. The formula employed for cosine similarity is as follows:

$$cosine\ similarity = \frac{A.B}{\|A\|\|B\|} \, , \tag{3}$$

Where A and B are the two vectors being compared, "." denotes the dot product of the two vectors, and ||A|| and ||B|| denote the Euclidean norms of the two vectors.

The resulting similarity score ranges from -1 to 1, where -1 indicates perfect dissimilarity and 1 indicates the perfect similarity of two documents.

3.3.3. Random Forests

Random forests are a widely employed machine learning algorithm, commonly applied in the realm of text mining tasks. The fundamental concept underlying a random forest is the construction of a group of decision trees that possess the capability to generate predictions concerning novel data instances. Random Forest, as a methodology, can be leveraged to effectively categorize text data into positive or negative sentiment classifications. The function of this algorithm entails the creation of an ensemble of decision trees, whereby each individual tree is trained on a random subset of the data, as well as a random subset of the features. The outcomes yielded by each tree are subsequently emerged to yield a definitive prediction.

The initial stage of using a random forest for text mining entails converting the textual data into a numerical format that the algorithm can use effectively. This conversion process is conventionally executed through the application of various techniques, such as bag-of-words or word embedding, which effectively represent the textual content as a vector comprising numerical values. Following the successful transformation of the data, the random forest algorithm can be trained using the annotated data and subsequently employed to generate accurate predictions for novel, unannotated data. One great thing about using a random forest for text mining is that it can handle high-dimensional data with a lot of different features well.

3.3.4. Logistic Regression

Logistic regression is a significant machine learning algorithm that can be used for text mining tasks. The basic idea behind logistic regression is to model the probability of positive or negative sentiment. In opinion mining, logistic regression is typically trained on a labelled dataset of text data, where each piece of text is labelled with its corresponding sentiment. The algorithm learns to identify patterns and features in the text data that are indicative of a particular sentiment, and logistic regression can be used for environmental sustainability text mining tasks. The equation of logistic regression can be expressed as:

$$p(y = 1|x) = \frac{1}{(1+\exp(-Z))}, \tag{4}$$

Where p(y = 1 | x) is the probability that the sentiment of the text x is positive (y = 1), given the input features of x and z is the linear combination of the input features of x and their corresponding weights learned during training:

$$z = w_0 + w_1x_1 + w_2x_2 + \ldots + w_nx_n, \tag{5}$$

Where w_0 is the intercept term and w_1 to w_n are the weights assigned to the input features of x_1 to x_n. x_1 to x_n are the input features extracted from the text x.

The sigmoid function is used to map the linear combination z to a probability value between 0 and 1:

$$\text{sigmoid}(z) = \frac{1}{(1+\exp(-Z))}. \tag{6}$$

The predicted sentiment label for the text x can then be determined based on the threshold value of the probability, which is typically set to 0.5. If p(y=1 | x) >= 0.5, then the predicted sentiment label is positive; otherwise, it is negative.

3.3.5. Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are commonly used in opinion mining for sentiment analysis tasks, where the goal is to determine the sentiment expressed in a piece of text. SVMs are effective in sentiment analysis tasks because they can learn to distinguish between positive and negative sentiments by identifying the important features or words that contribute to each sentiment. The formula for SVM classification can be represented as:

$$f(x) = sign(w \,.\, x + b), \tag{7}$$

Where x is the input text, w is the weight vector, b is the bias term, and sign () is the sign function that returns +1 for positive sentiment and -1 for negative sentiment.

The weight vector and bias term are learned during training using a set of labelled data, where each data point is a text example with its corresponding sentiment label. The goal of the training process is to find the optimal weight vector and bias term that maximize the margin between the positive and negative classes, while minimizing the misclassification error.

3.3.6. Accuracy

In sentiment analysis, classification accuracy refers to the percentage of correctly classified opinions or sentiments as positive or negative. It is a measure of the effectiveness of the opinion mining algorithm in accurately identifying the sentiment of the given text. The formula for calculating the accuracy of a classification model is as follows:

$$\text{Accuracy} = \frac{(TP+TN)}{TP+TN+FP+FN}, \tag{8}$$

Where TP (true positive) is the number of positive and TN (true negative) is the number of negative. Similarly, FP (false positive) is the number of negative opinions that were incorrectly identified as positive and FN (false negative) is the number of positive opinions that were incorrectly identified as negative.

The accuracy for positive sentiment classification can be calculated using the following formula:

$$\text{Positive Accuracy} = \frac{(TP)}{P}. \tag{9}$$

The accuracy for negative sentiment classification can be calculated using the following formula:

$$\text{Nagative Accuracy} = \frac{(TN)}{(N-P)}. \tag{10}$$

Let's assume that dataset has N opinions, out of which there are P positive opinions and N-P negative opinions.

3.3.7. Precision, Recall, and F-Measure

In text mining, precision, recall, and F-measure are common ways to measure how well machine learning models or natural language processing techniques work on text data [29] that has to do with sustainability.

Precision measures the accuracy of the positive predictions made by the model. In the context of environmental sustainability, precision can be used to evaluate the accuracy of the model in identifying relevant environmental terms or concepts in text data by measuring how many of the positive predictions are actually true positives, which is calculated as follows:

$$precision = \frac{true\,positive}{true\,positive + false\,positive}. \quad (11)$$

Recall measures the completeness of the positive predictions made by the model. It measures how many of the actual positives were correctly identified by the model, and it is calculated as:

$$recall = \frac{true\,positive}{true\,positive + false\,negative}. \quad (12)$$

F-measure is a weighted average of precision and recall that combines the two metrics into a single score. It is calculated as the harmonic mean of precision and recall. The F-measure is useful in situations where both precision and recall are important, and it can be used to compare the performance of different models or techniques on the same dataset. It is calculated as:

$$F - Measure = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)}. \quad (13)$$

4. Results and Discussion

The result gained from the experiment by using logistic regression, random forest, and support vector machine algorithms is shown in Table 1. The data from this project was collected from social networks and accuracy, precision, recall, and F-measure were used to evaluate the effectiveness of classification models.

Table 1.
Classification accuracy comparison.

Topic	Logistic regression	Random forest	Support vector machines
Positive	0.78	0.74	0.89
Negative	0.85	0.82	0.91

Table 1 presents the classification accuracy comparison among the three algorithms with TF/IDF and cosine similarity. Support vector machines had the highest accuracy, while random forest had the lowest.

Table 2.
The precision, recall and F-measure results.

Machine learning techniques	Testing performance		
	Precision	Recall	F1-score
Logistic regression	83.87	84.54	82.25
Random forest	85.59	82.14	79.75
Support vector machines	89.47	89.52	89.64

Table 2 shows the results of the precision, recall, and F-measure of machine learning approaches. The lowest precision of the models is presented by logistic regression 83.87% and a recall of 84.54% in the context of logistic regression means that the model correctly identified approximately 84.54% of all actual positive instances in the dataset. The logistic regression model achieved an F1-score of 82.25%, which indicates a good balance between precision and recall, and this suggests that the model has a relatively low rate of both false positives and false negatives. Also, a precision of 85.59% for a Random Forest model means that approximately 85.59% of the instances predicted as positive by the Random Forest model were indeed positive. However, An F1-score of 79.75% for a Random Forest model is the lowest F1-score. Support vector machine algorithm shows the best performance results in accuracy, precision, recall, and f1-score compared to other algorithms (89.47% for precision, 89.52% for recall, and 89.64% for f1-score).

In the context of machine learning, precision, recall, and F-measure are evaluation metrics used to assess the performance of a classification or information retrieval system. Precision measures the accuracy of the positive predictions, and recall measures the ability of the model to capture all the actual positives. Furthermore, the F1-score combines precision and recall to provide a single metric that considers both false positives and false negatives. From this research, Support Vector Machines (SVM) can be a suitable machine learning algorithm to assess the impact of social networks on the awareness of environmental sustainability. This interpretability can help to understand what aspects of social network activity influence environmental awareness.

Sentiment analysis can offer valuable insights into environmental sustainability awareness by analyzing the language, opinions, and emotions expressed in online content, such as social media posts, articles, and forums. From these results, sentiment analysis can gauge how people feel about environmental sustainability initiatives, policies, or events. Positive sentiment might indicate widespread support and enthusiasm for sustainability efforts, while negative sentiment could signal dissatisfaction or skepticism. Also, by analyzing sentiments over time, it becomes possible to identify trends in public awareness and concern for environmental issues. For instance, an increase in positive sentiment might suggest growing interest in or awareness of sustainability-related topics. Understanding sentiment can also shed light on the influence of key figures or influencers in driving conversations around environmental sustainability. Positive sentiment among influential figures could amplify awareness and engagement in sustainability efforts. Negative sentiment can provide constructive feedback. By examining why certain sustainability initiatives are perceived negatively, organizations can refine their approaches, address misconceptions, or adapt strategies to better resonate with their audience. By leveraging sentiment analysis, organizations, policymakers, and businesses can gain deeper insights into public attitudes and perceptions towards environmental sustainability. These insights can inform strategies to enhance awareness, engagement, and the effectiveness of sustainability initiatives.

5. Conclusion

In this particular implementation, the application of machine learning serves as a highly effective tool for the purpose of sentiment analysis. This process involves the identification and analysis of individuals' opinions, attitudes, and emotions with regards to the topic of environmental sustainability. By conducting sentiment analysis, one gains a deeper understanding of people's perceptions and attitudes towards environmental issues and sustainability initiatives. Moreover, the classification of sentiments into positive or negative categories further aids in the analysis of textual data. Opinion mining tasks greatly benefit from the utilization of powerful supervised learning algorithms such as logistic regression, decision trees, or support vector machines. Prior to the application of these algorithms, it was common practice to preprocess and analyze text data using techniques such as TF/IDF and cosine similarity. Sentiment analysis proves to be a valuable approach to comprehending individuals' attitudes and opinions regarding environmental sustainability. The analysis involves the examination of language data, such as social media posts or customer reviews, in order to determine whether the sentiments expressed are positive or negative.

In conclusion, sentiment analysis can be a useful tool for tracking progress, identifying influencers, monitoring public perceptions of environmental sustainability, and enhancing communication. Organizations can more effectively engage and inspire the public to adopt sustainable habits by utilizing sentiment analysis.

Funding:
This research is supported by Suan Sunandha Rajabhat University, Thailand (Grant number: PD03-6602007).

Institutional Review Board Statement:
Not applicable.

Transparency:
The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Competing Interests:
The authors declare that they have no competing interests.

Authors' Contributions:
All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

References

[1] Department of statistics in the sumy region, "Department of statistics in the sumy region," Retrieved: http://sumy.ukrstat.gov.ua/?menu=105&level=3/. [Accessed 31-03-2015], 2015.
[2] World Wide Fund for Nature, "World wide fund for nature," Retrieved: https://wwf.panda.org/wwf_news/?337389/Nothing-natural-about-natures-steep-decline-WWF-report-reveals-staggering-extent-of-human-impact-on-planet. [Accessed 16-02-2023], 2023.
[3] Natural-Resource Use and Environmental Impacts, "Natural-resource use and environmental impacts," Retrieved: https://www.oneplanetnetwork.org/SDG-12/natural-resource-use-environmental-impacts. [Accessed 26-01-2023], 2023.
[4] Climate change management and coordination, "Climate change management and coordination," Retrieved: https://climate.onep.go.th/en_US/knowledge/thailand-vs-climate-change/ [Accessed 29-01-2023], 2023.
[5] CO2 Emissions by Country, "CO2 emissions by country," Retrieved: https://www.worldometers.info/co2-emissions/co2-emissions-by-country/. [Accessed 29-01-2023], 2023.
[6] Climate Change Performance Index, "Climate change performance index," Retrieved: https://ccpi.org/country/tha/. [Accessed 29-01-2023], 2023.
[7] A. Ligthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: A tertiary study," *Artificial Intelligence Review*, vol. 54, no. 7, pp. 4997–5053, 2021. https://doi.org/10.1007/s10462-021-09973-3
[8] M. J. Awan *et al.*, "Fake news data exploration and analytics," *Electronics*, vol. 10, no. 19, p. 2326, 2021. https://doi.org/10.3390/electronics10192326
[9] K. Kularbphettong, "The awareness of environment conservation based on opinion data mining from social media," *GEOMATE Journal*, vol. 17, no. 61, pp. 74-79, 2019. https://doi.org/10.21660/2019.61.4700
[10] N. F. Baarir and A. Djeffal, "Fake news detection using machine learning," in *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-Being (IHSH)*, 2021, pp. 125-130.
[11] A. Hadeer, T. Issa, and S. Sherif, "Detection of online fake news using n-gram analysis and machine learning techniques," presented at the International Conference on Intelligent Secure and Dependable Systems in Distributed and Cloud Environments, pp. 127-138, 2017.
[12] N. R. Ramadhanti and S. Mariyah, "Document similarity detection using indonesian language word2vec model," presented at the 3 rd International Conference on Informatics and Computational Sciences (ICICoS), pp. 1-6, 2019.
[13] J. Singh and P. Tripathi, "Sentiment analysis of twitter data by making use of svm, random forest and decision tree algorithm," presented at the 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2021, pp. 193-198,, 2021.

[14] S. Clark and R. Wicentwoski, "Swatcs: Combining simple classifiers with estimated accuracy," in *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, 2013, pp. 425-429.
[15] B. Bahrawi, "Sentiment analysis using random forest algorithm-online social media based," *Journal of Information Technology and Its Utilization*, vol. 2, no. 2, pp. 29-33, 2019. https://doi.org/10.30818/jitu.2.2.2695
[16] P. Karthika, M. R., and R. Manoranjithem, "Sentiment analysis of social media network using random forest algorithm," presented at the 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Tamilnadu, India, 2019, pp. 1-5, 2019.
[17] S. Bhattacharjee, A. Das, U. Bhattacharya, S. K. Parui, and S. Roy, "Sentiment analysis using cosine similarity measure," presented at the 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS), Kolkata, India, 2015, pp. 27-32, 2015.
[18] S. Gurdeep, T. Praveen, K. S. Rohit, and S. Ram, "Sentiment analysis in social media using machine learning technique with r language," *Journal of Chemical and Phamaceutical Sciences*, vol. 6, pp. 124-128, 2017.
[19] A. Bhatt, A. Patel, H. Chheda, and K. Gawande, "Amazon review classification and sentiment analysis," *International Journal of Computer Science and Information Technologies*, vol. 6, no. 6, pp. 5107-5110, 2015.
[20] U. R. Babu and N. Reddy, "Sentiment analysis of reviews for e-shopping websites," *International Journal of Computational Science*, vol. 6, no. 1, p. 19966, 2017. https://doi.org/10.18535/ijecs/v6i1.20
[21] B. Deewattananon and U. Sammapun, "Analyzing user reviews in Thai language toward aspects in mobile applications," in *2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2017, pp. 1-6.
[22] A. Tweety, "Internet's most loved twitterapi python library," Retrieved: http://www.tweepy.org/. [Accessed 29-01-2023], 2023.
[23] X. Tao, X. Zhou, J. Zhang, and J. Yong, "Sentiment analysis for depression detection on social networks," in *Advanced Data Mining and Applications: 12th International Conference, ADMA 2016, Gold Coast, QLD, Australia, December 12-15, 2016, Proceedings 12*, 2016, pp. 807-810.
[24] Z. Jianqiang and G. Xiaolin, "Comparison research on text pre-processing methods on twitter sentiment analysis," *IEEE Access*, vol. 5, pp. 2870-2879, 2017.
[25] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis," *Expert Systems with Applications*, vol. 110, pp. 298-310, 2018. https://doi.org/10.1016/j.eswa.2018.06.022
[26] Thai Stop Word, Retrieved: Sign in to GitHub · GitHub. 2023.
[27] PythaiNLP library, "Pythainlp 5.0.1," Retrieved: https://pypi.org/project/pythainlp/. 2023.
[28] A. Dey, M. Jenamani, and J. J. Thakkar, "Lexical TF-IDF: An n-gram feature space for cross-domain classification of sentiment reviews," in *International Conference on Pattern Recognition and Machine Intelligence*, 2017, pp. 380-386.
[29] P. Max Bramer, "Principles of data mining." London: Springer, 2007, pp. 173-176.