**MEDIA AND SOCIAL NETWORKS APPROACH FOR ENVIRONMENTAL CONCERNS ASSESSMENT, A CASE STUDY VIETNAM**

**Ngoc Anh Ngo1, Mai Dung Nguyen1**

1Faculty of Information Technology, Hanoi University of Mining and Geology, Hanoi, Vietnam

 **ABSTRACT**

**The environment of developing country like Vietnam often facing many troubles while it pays attention to the economic development. Data generated from social media can play important roles in raising awareness and informing policy decisions about environmental pollutions in Vietnam. Social media platforms have a wide reach, allowing for monitoring of environmental issues in remote or less accessible areas where classical monitoring methods may be logistically challenging or expensive to implement. They also provide near-real-time information on environmental events. This allows for faster response times to emerging issues compared to traditional data collection methods which may have delays in reporting.**

**Keywords:** Sentimentt Analysis, Tensorflow, Social Media, Environment, Machine learning.

1. **INTRODUCTION**

The environment of developing country like Vietnam often facing many troubles while it pays attention to the economic development. Rapid urbanization, industrialization, and a high number of motorized vehicles contribute to elevated levels of air pollution. This poses significant risks to public health and requires effective air quality management measures. Industrial discharge, improper waste disposal, and inadequate wastewater treatment contribute to water pollution in rivers and lakes. Inadequate waste collection, recycling, and disposal systems lead to visible litter and environmental degradation. Urban areas experience significant levels of noise pollution, affecting residents' quality of life. Addressing these urgent environmental issues requires concerted efforts from government authorities, urban planners, environmental agencies, and the community. Implementing sustainable policies, investing in green technologies, and raising public awareness are all key components of mitigating these challenges in big cities in Vietnam.

As of 2022, there were approximately 69 million Facebook users in Vietnam. This number may have continued to grow since then. Data generated from social media can play important roles in raising awareness and informing policy decisions about environmental pollutions in Vietnam. Social media platforms such as Facebook, Twitter and Instagram can provide rich sources of data on public opinions, sentiments and behaviors related to environmental issues. For example, social media data can help identify the most prevalent and urgent environmental problems that affect the Vietnamese people, such as air pollution, water pollution and waste management. Social media data can also help monitor the impacts of environmental policies and interventions, such as the effectiveness of banning single-use plastics, promoting renewable energy sources and encouraging green lifestyles. Indeed, numerous cases of environmental pollution in Vietnam have initially come to light through social networks and news outlets, subsequently drawing the attention of both authorities and communities, e.g., massive fish death in the central cost in 2016, rafinery oil leak in Vung Tau in 2018. These incidents highlight the importance of vigilance, reporting, and public awareness in addressing environmental pollution. Social media and news outlets play a crucial role in disseminating information, mobilizing public response, and holding responsible parties accountable for their actions. They also contribute to greater transparency and awareness of environmental issues in Vietnam.

Machine learning has been increasingly used in sentiment analysis for environmental issues and natural hazards. Here are some examples:A study examined local governments' use of social media in disclosing environmental actions/plans/information to improve accountability to citizens [1]. The researchers computed the sentiment of citizens' comments using two different approaches: one based on a lexicon dictionary and the other based on convolutional neural networks [1]. Another study performed a comparative evaluation of various sentiment analysis approaches for climate change tweets [2]. The study used seven lexicon-based approaches and three machine learning classifiers, and found that the hybrid method outperformed the other two approaches. A framework was developed to analyze users’ sentiments on Twitter on natural disasters using a hybrid of machine learning, statistical modeling, and lexicon-based approach [5]. The study drew insights from tweets about Kerala’s 2018 natural disasters in India. Another study tested the accuracy of the pre-trained sentiment analysis model developed by MonkeyLearn using text data related to the emergency response and early recovery phase of the earthquakes that struck Albania on November 26, 2019[6].

These studies demonstrate the potential of machine learning in sentiment analysis for environmental issues and natural hazards, providing valuable insights for decision-making.

1. **METHODOLOGY**

The proposed methodology is illustrated in Figure . Information from social media platforms like Facebook and Twitter is automatically retrieved and saved in a database. This content is subsequently pre-processed to generate both a training set and a testing set. The resulting model, once tested, is employed to forecast newly gathered data from social networks and news sources. The forecasted results are then presented visually on a map.



**Figure 1:** Sentiment analyzing process with deep learning approach

* Data Collection: Collect data from social media platforms and news websites. You can use APIs provided by these platforms or web scraping tools to collect data. Make sure to respect the terms of service of each platform.
* Data Preprocessing: Removing irrelevant or duplicate posts, comments, or hashtags that do not relate to the environmental issues of interest. Extracting relevant information from the text, such as keywords, topics, sentiments, opinions, or emotions. Normalizing the text by correcting spelling, grammar, punctuation, or slang. Converting the text into numerical or categorical values that can be used for statistical or machine learning methods. Merging or joining the data from different platforms or accounts that belong to the same user or entity.
* Content Analysis: Analyze the content of the posts or articles using Natural Language Processing (NLP) techniques. This could involve sentiment analysis to understand public opinion, topic modeling to identify common themes, or named entity recognition to identify specific entities related to environmental issues.
* Assess Environmental Issues: Based on the results of your content analysis, assess the environmental issues you defined in step 1. This could involve correlating public sentiment with real-world data on these issues, or identifying trends in how these issues are discussed over time.
* Data Visualization: Visualize your findings using appropriate charts and graphs. This could include word clouds for most common words, bar charts for sentiment scores, or line graphs for trends over time.
1. **IMPLEMENTATION**
	1. **Crawling tool**

The paper suggests employing the ParseHub tool for automated data collection from social networks. ParseHub is a web scraping tool equipped with a user-friendly point-and-click interface, capable of extracting data from any website. The process of crawling is outlined in the subsequent steps:

1. Create a project in ParseHub and select the social network website that popular in Vietnam to scrape data from.

2. Define the data elements that to extract, such as posts, comments, likes, shares, etc. using the selectors and commands in ParseHub to navigate the website and capture the data.

3. Save the crawl data to the database.

* 1. **Sentiment Analysis for Environmental Issues**

This method represents a significant advancement in sentiment analysis by employing TensorFlow's robust sequential modeling capabilities. The proposed algorithm aims to provide a more nuanced and accurate understanding of public sentiment towards environmental issues that includes the following steps:

1. Data Preparation: The texts are tokenized, meaning they are split into individual words. Each word is then mapped to a unique integer. The texts are then transformed into sequences of these integers. If the texts have different lengths, they are padded so that they all have the same length.
2. Loading Pre-Trained Word Embeddings: The GloVe pre-trained word embeddings are loaded. These embeddings map each word in the vocabulary to a 100-dimensional vector.
3. Preparing the Embedding Matrix: An embedding matrix is created where the i-th row gives the embedding vector for the word represented by integer i. This matrix is used to initialize the weights of the Embedding layer in the model.
4. Building the Model: The model is built as a Sequential model, which means it consists of a linear stack of layers. The first layer is an Embedding layer that turns positive integers (indexes) into dense vectors of fixed size (the embeddings). It is configured to be non-trainable, which means its weights (the embedding matrix) will not be updated during training. The output from the Embedding layer is then averaged for each sequence, resulting in a fixed-length output vector for each sequence regardless of its length. This fixed-length output vector is piped through a fully connected (Dense) layer with 16 hidden units. The last layer is also a Dense layer with 1 output unit and a sigmoid activation function, which will output the predicted sentiment (0 for negative, 1 for positive).
5. Compiling the Model: The model needs to be compiled before it can be trained. During compilation, the loss function, optimizer and metrics are specified. The loss function is binary cross-entropy which is suitable for binary classification problems like this one. The optimizer is Adam, which is an algorithm for first-order gradient-based optimization of stochastic objective functions.
6. Training the Model: The model is trained on the data for a specified number of epochs or exposures to the training dataset.

The snippets below show the code of the proposed model

*tokenizer = Tokenizer(num\_words=vocab\_size, oov\_token=oov\_tok)*

*tokenizer.fit\_on\_texts(training\_sentences)*

*training\_sequences = tokenizer.texts\_to\_sequences(training\_sentences)*

*training\_padded = pad\_sequences(training\_sequences, maxlen=max\_length, padding=padding\_type, truncating=trunc\_type)*

*testing\_sequences = tokenizer.texts\_to\_sequences(testing\_sentences)*

*testing\_padded = pad\_sequences(testing\_sequences, maxlen=max\_length, padding=padding\_type, truncating=trunc\_type)*

*model = tf.keras.Sequential([*

 *tf.keras.layers.Embedding(vocab\_size, embedding\_dim, input\_length=max\_length),*

 *tf.keras.layers.GlobalAveragePooling1D(),*

 *tf.keras.layers.Dense(24, activation='relu'),*

 *tf.keras.layers.Dense(1, activation='sigmoid')*

*])*

*model.compile(loss='binary\_crossentropy',optimizer='adam',metrics=['accuracy'])*

*num\_epochs = 100*

*history = model.fit(training\_padded, training\_labels, epochs=num\_epochs, validation\_data=(testing\_padded, testing\_labels), verbose=2)*

1. **RESULTS AND DISCUSSION**



**Figure 2. Evaluating the model**

The results of utilizing the sequential model for sentiment analysis on the provided dataset are quite promising. The dataset comprises 964 instances, consisting of headlines extracted from various sources including Facebook and the web. Among these instances, 483 were classified as having a positive sentiment, while 408 were labeled as negative. For evaluating the model's performance, a separate test set of 100 instances was employed. The training phase yielded an impressive accuracy of 98.40%, indicating the model's adeptness in learning from the training data. Subsequently, when tested on the dedicated test set, the model demonstrated a commendable accuracy of 92%, signifying its ability to generalize well to new, unseen data. These results suggest that the sequential model exhibits strong predictive capabilities in discerning sentiment from environmental headlines.

1. **CONCLUSION**

This paper presents a methodology for sentiment analysis in the context of environmental discourse in Vietnam, harnessing the capabilities of sequential models within the TensorFlow framework. The results obtained from the extensive evaluation on a dataset comprising 964 instances, sourced from platforms like Facebook and the web, are highly encouraging. With a training accuracy of 98.40% and a test accuracy of 92%, the model demonstrates a remarkable proficiency in discerning sentiment, showcasing its effectiveness in real-world applications. This innovative approach not only enhances our understanding of public sentiment towards environmental issues but also holds immense potential for informing evidence-based policy-making and targeted interventions in the field of environmental conservation. The success of this methodology opens up avenues for further research and applications in sentiment analysis across various domains, ultimately contributing to more informed and effective decision-making processes.

**ACKNOWLEDGEMENTS**

The authors can acknowledge professor, friend or family member who help in research work in this section.

1. **REFERENCES**
2. Rocca, L., Giacomini, D. and Zola, P. (2021), "Environmental disclosure and sentiment analysis: state of the art and opportunities for public-sector organisations", Meditari Accountancy Research, Vol. 29 No. 3, pp. 617-646. https://doi.org/10.1108/MEDAR-09-2019-0563
3. Mohamad Sham, N.; Mohamed, A. Climate Change Sentiment Analysis Using Lexicon, Machine Learning and Hybrid Approaches. Sustainability 2022, 14, 4723. https://doi.org/10.3390/su14084723
4. Mendon, S., Dutta, P., Behl, A. et al. A Hybrid Approach of Machine Learning and Lexicons to Sentiment Analysis: Enhanced Insights from Twitter Data of Natural Disasters. Inf Syst Front 23, 1145–1168 (2021). https://doi.org/10.1007/s10796-021-10107-x
5. Contreras, D., Wilkinson, S., Alterman, E. et al. Accuracy of a pre-trained sentiment analysis (SA) classification model on tweets related to emergency response and early recovery assessment: the case of 2019 Albanian earthquake. Nat Hazards 113, 403–421 (2022). https://doi.org/10.1007/s11069-022-05307-w
6. Yadav, A., Vishwakarma, D.K. Sentiment analysis using deep learning architectures: a review. Artif Intell Rev 53, 4335–4385 (2020). https://doi.org/10.1007/s10462-019-09794-5.
7. Lei Zhang, Shuai Wang, & Bing Liu (2018). Deep Learning for Sentiment Analysis : A Survey. CoRR, abs/1801.07883..
8. Yadav, A., Vishwakarma, D.K. Sentiment analysis using deep learning architectures: a review. Artif Intell Rev 53, 4335–4385 (2020). https://doi.org/10.1007/s10462-019-09794-5