CMSI: Carbon Market Sentiment Index with AI Text Analytics

Min-Yuh Day*
Graduate Institute of Information Management
National Taipei University
New Taipei City, Taiwan
myday@gm.ntpu.edu.tw

Chia-Tung Tsai
Graduate Institute of Information Management
National Taipei University
New Taipei City, Taiwan
s711136106@gm.ntpu.edu.tw

***Abstract*—Climate change is an increasing environment concern, carbon markets are received attention because they can reduce emissions by carbon trading. With the rising in sustainable finance demand, there need some innovative tools to study investors' attitude. Although there has been an increasing amount of literature on market indicator and sentiment analysis, there is a limited research focus on carbon market. Therefore, this research purpose is to develop carbon market sentiment indicator. The research method contains three mainly phase. By using pretrained models like BERT, GPT, and so on, the research can obtain sentiment analysis results from news and social media data. Then, taking a specific carbon market as the object, and the selected sentiment indicators, the carbon market response for a period can be obtained at last. Due to complexity and the lack of transparency in the institutional and financial infrastructure for carbon market transactions, there has the problem of market volatility. The contribution of research is that provides a comprehensive sentiment index for carbon market. By this way, the research expects to provide valuable insights into sustainable finance and help investors to make better decision.**

Keywords—carbon market, net-zero emissions, sustainable finance, social media, sentiment analysis

I. INTRODUCTION

Since climate change impacts on the economy, environment, and society, several countries are beginning to focus on this issue and find solution to cope with it. Carbon emission reductions have gotten more attention around the world [5]. Many countries regard controlling greenhouse gas emissions and developing a low-carbon economy as a their same objective [8]. However, their success depends on market participants' support. Therefore, understanding how market participants feel about carbon markets is crucial [11].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ASONAM '23, November 6–9, 2023, Kusadasi, Turkey
© 2023 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0409-3/23/11…$15.00
https://doi.org/10.1145/3625007.3627516

Previous research has used sentiment analysis towards carbon market. Wen et al. [14] indicated that market sentiment is a driving factor in carbon price dynamics. Xiang et al. [17] employed sentiment analysis to investigate the perspective of Twitter users on China's carbon policy. Although these researchers have studied on sentiment analysis and carbon market, there is little published paper on construct sentiment index on carbon market. This research attempts to fill the gap and increase the existing literature on constructing sentiment indicator in carbon market.

Previous research about sentiment indicator in carbon market has used methods like questionnaire surveys to understand how market participants feel about carbon markets[20]. However, these methods can be time-consuming, expensive, and biased. Artificial intelligence technology is a new research method that can analyze large amounts of text data quickly and accurately, providing more precise insights into market sentiments and attitudes.

This study will use text analysis to get sentiment scores from social media texts and news articles. Then combine the carbon market prices, including China, the EU, Singapore, and Taiwan carbon market to obtain a volatility score. Finally, the comprehensive index CMSI is obtained through calculation.

By analyzing news and social media related to the carbon market, we can better understand market participants' emotions and attitudes. This can increase the chance of successful implementation of the carbon market.

The remainder of this paper is structed as follows. Section II provides an overview of the related work. Section III illustrates the research methodology and proposed system architecture. Section IV shows the data analysis and discussion. Finally, Section V presents the paper's conclusion.

II. RELATED WORK

A. Social Media

With the development of technology, Internet changed the way people express their views [21]. Using sentiment analysis on social media platforms like YouTube and Facebook has a variety of applications [22].

Since Twitter text has abundant content, as well as easy to access for everyone. There are millions of tweets on platforms to discuss different issues in every day. This shows social networks are developing rapidly into a kind of data source [23].

Peng et al. [2] proposed a useful approach that can be used to analyze the depression level of people on Twitter. This research helps people comprehend the effects of COVID-19 on people's psychological health. Chatsiou [4] shown that the

TABLE I. USING SENTIMENT ANALYSIS IN DIFFERENT SOCIAL MEDIA

Social Media	Ref.	Method	Target	Accuracy (%)
Twitter	Peng et al. [2]	BERT	The exploitation of the largest Twitter English depression dataset	76.5
	Chatsiou [4]	CNN+BERT	Auto-assign sentences for the corpus of the COVID-19 press briefing	68.7
Facebook	Biswas et al. [7]	BERT	Predict user sentiment from Facebook reviews of online food delivery services	92.9
	Kang et al. [10]	BERT	Understand the brand image of Malaysian airline companies	86.0
Weibo	Figueira et al. [13]	BERT+BiLSTM+Att	Extract sentiment features and conduct comparative experiment during the COVID-19	89.7
	Li and Li [16]	GNN-LSTM	Capture the significant feature of sentiment detection	95.3

performance of CNN combined with transformer-based model like BERT is better. By this way, it can classify sentences automatically. Biswas et al. [7] collected users' reviews from Facebook. Then comprehended and categorized the users' feedback for food delivery services. It is vital for business to consider the sentiment of their customers. This helps them to do better decision in the future by analyzing these reviews. Kang et al. [10] studied about using Facebook reviews to understand Malaysian airline companies social status, reputation, and brand image. Table I provides recent research on applying sentiment analysis to social media texts.

Overall, this research uses social media resources for sentiment analysis and provides a valuable source of data that can help the government monitor public sentiment in real-time and examine potential factors associated with articles on the carbon market.

B. Sustainable Finance

The definition of sustainable finance is very extensive, Kumar et al. [24] suggested it should contain all activities and factors that make finance sustainable and improve sustainable development. Various ways such as climate finance, green bond and carbon emissions can be included under the definition of sustainable finance. Kumar et al. [24] conducted literature review on sustainable finance. They found out a seminal literature by Ferris and Rykaczewski [25]. Al Muhairi and Nobanee [26] explained sustainable financial management enhances the capability of environment, social and governance. It provides a procedure for investments for eliminating climate change, reducing carbon emissions, and rising energy efficiency.

Carbon market encourages firms and organizations to reduce carbon emissions through trading to achieve emission reduction goals. Net-zero emissions is to obtain a balance between carbon emissions and absorption. Sustainable finance includes above mentioned concepts. Therefore, we will introduce them.

The key determining factor in the operation of an emissions trading system is the price of carbon. If its fluctuations are too severe, they will have a direct impact on achieving emission reduction targets as well as societal sustainable development [27].

EU Emissions Trading System and Chicago Climate Exchange are two mainly systems for carbon emissions trading.

The largest carbon emission trading market is the European Union Emission Trading System (EU ETS). Furthermore, it has a demonstrable effect on other carbon markets [28]. The EU ETS is important because it assists the European Union mitigate the effects of carbon emissions on global warming [29]. It defines carbon emission quotas and regulates carbon dioxide emissions from different industries. Using market mechanisms for carbon emission rights trading to solve the issue between environmental concerns and low-carbon economic development.

It utilizes market mechanisms to proceed carbon emission trading and solve low-carbon economic development. The EU ETS is international carbon trading system. It makes different countries to achieve emission reduction commitments by using a common mechanism [30].

Furthermore, China's carbon market is an significant tool to reduce carbon emissions [31]. China carbon market will help reduce global greenhouse gas emissions and improve the development of the global carbon emission reduction. In addition, carbon market also demonstrates the Chinese government's commitment and action to address climate change. It has a great significance to global climate governance.

Singapore plans to launch Climate Impact X, a global exchange and marketplace for transparent, high-quality, and high-integrity carbon offset credits [32].

Net-zero emissions refers to balance between the amount of greenhouse gas produced and the amount removed from the atmosphere. It can be accomplished through a combination of emission reduction and emission removal.

When decision makers formulate policies, such as halting carbon emissions, focusing on public policy and resource allocation on development. It can assist to bring carbon emissions down to zero [32]. Liu and Lin [33] applied time series forecasting with deep learning model to study the effectiveness between COVID-19 virus spreading and net-zero target of 2050.

C. Sentiment Analysis

Sentiment analysis extracts and analyses people's opinion and detect the sentiment polarity from text [34]. Most data in social media platforms is unstructured. This makes it harder

TABLE II. OVERVIEW OF SENTIMENT ANALYSIS IN DIFFERENT DOMAIN

Author(s)	Domain	Method	Dataset	Evaluation	Performance (%)
Liu et al. [1]	Business	BERT-BiGRU	COAE2014-task4	Accuracy, F1 Score	89.0, 88.6
Xiao and Luo [3]	Business	BERT-LSTM	Semeval2014	Accuracy, F1 Score	82.2, 72.5
Edalati et al. [6]	Education	BERT	Coursera	Precision, Recall, F1 Score	91.1, 92.3, 92.0
Su and Peng [9]	Education	RoBERTa	China's MOOC Ke reviews	Accuracy, F1 Score	93.4, 94.5
Li et al. [12]	Government	BERT + BiLSTM + Att	Weibo	Accuracy, Percision, Recall, F1 Score	89.7, 90.9, 87.8, 89.3
Cai et al. [15]	Government	BERT-BiLSTM	Chinese Internet Review	Recall, Accuracy, Precision, F1 Score	70.7, 86.2, 57.7, 63.5
Aygün et al. [18]	Health	mBERT-base	Twitter	Accuracy, Recall, F1 Score	86.0, 83.0, 84.0
Chandra and Krishna [19]	Health	BERT	Senwave COVID-19 sentiment dataset	F1 macro, F1 micro	53.0, 58.7

to assess and extract meaningful information from these data [35].

Sentiment analysis can achieve various goals like observing public attitude, understanding customer satisfaction and so on [36]. Sentiment analysis classifies opinions into positive, negative, or neutral [21].

Most of the literature categorizes approaches into three types: machine learning approaches, lexicon-based approaches, and hybrid approaches.

Machine learning can classify sentiments. Sentiment analysis can used in two methods: supervised machine learning and lexicon-based unsupervised learning [37]. By using supervised machine learning, there are some common algorithms including Naïve Bayes, Support Vector Machine, Logistic Regression, Decision Tree, Maximum Entropy, and K-Nearest Neighbours. By using deep learning, some common neural network models such as CNN, RNN, and DNN. They can learn features from the dataset by themselves.

Lexicon-based approach employs an emotion lexicon to describe the polarity, such as positive, negative, and neutral. Researchers compile sentiment word lists to create a sentiment lexicon. According to positive and negative indicator, they can determine polarity scores [38]. Compared with the algorithm of machine learning, this approach is easier to comprehend [39]. This approach has two categories: dictionary-based approach and corpus-based approach. For dictionary-based, the advantage of it is not require training data. However, it cannot find opinion terms that are not contained in the lexicon [38].

Hybrid approach combines machine learning and lexicon-based approaches. The high accuracy in machine learning and stability in lexicon-based approach make some researchers utilized this approach to analysis [40]. Elshakankery and Ahmed [41] combined machine learning and lexicon-based approach to understand sentiment polarities from Twitter.

To date, several studies have extracted people's opinion and perspective by using sentiment analysis. We reviewed some studies in different domains that are as follows:

In the areas of business and e-commerce, there has many advantages for applying sentiment analysis. For instance, firms can utilize the results of sentiment analysis to improve their products, analyze consumers' opinions, or provide innovative market strategies [42]. To analyze e-commerce product reviews, Liu et al. [1] proposed the BERT-BiGRU-Softmax model. Xiao and Luo [3] purposed ASBABM, an aspect-level sentiment analysis model based on BERT and multi-level attention mechanisms.

In terms of the educational domain, student feedback is critical in evaluating and analyzing learning platforms, teaching, and courses [43]. Through an attention mechanism, the Transformer model can learn the student's feedback. It can also categorize unlabeled comments to predict their emotion [44]. Zhou and Ye [45] did a statistical analysis on 41 review papers published between 2010 and 2020. The amount of sentiment analysis research has been published in journals has significantly increased. It illustrates that sentiment analysis is increasingly vital in education domain.

In terms of the government domain, sentiment analysis is a helpful tool to monitor public emotion [46]. Cai et al. [15] understand the energy market trend for investors and consumers by using sentiment analysis.

In the medical domain, Aygün et al. [18] applied aspect-based sentiment analysis for the United States, United Kingdom, Canada, Turkey, France, Germany, Spain, and Italy during COVID19 period to demonstrate twitter users' attitudes toward vaccination and vaccine types. Kastrati et al. [47] used sentiment analysis on Facebook posts about the spread of the virus.

During the increase in COVID-19 infections, Chandra and Krishna [19] used deep learning models to do sentiment analysis. The paper revealed optimism, fear, and uncertainty.

In the finance domain, specialized language and unlabeled data make sentiment analysis become a challenging task. FinBERT is a financial-based fine-tuning of the BERT language representation model [48].

D. Pre-trained Language Models in Sentiment Analysis

Google proposed the Transformer model in 2017, which utilizes attention mechanisms to process input sequences instead of traditional recurrent neural networks or convolutional neural networks [49]. In recent years, transformer model plays an important role in natural language processing. Transformer focus on the relationship between all words in a sentence instead of their respective position [50].

Transformer based on an encoder-decoder model. The encoder is used to analysis a sequence of input data and obtaining an encoding. The decoder is used to obtain an output from the encoding [51]. This architecture can be used in different tasks, such as question answering, summarization, and so on.

BERT is a transformer-based model and just use a transformer encoder [52]. This model is the first unsupervised and bidirectional system for performing natural language processing tasks [50].

The BERT model can be used in two steps: pre-training and fine-tuning [53].The BERT model's pre-training task has two tasks: Masked Language Model (MLM) and Next Sentence Prediction (NSP) [53]. MLM is intended to assist the model in understanding the contextual relationships in sentences and predicting the masked words. 15% of the words in the training data are chosen at random to be masked and replaced with [MASK] tags or random words. The model then predicts the correct or random words for these tags [15]. Traditional language models only consider the previous words; however, in MLM, the model must use contextual information to infer the masked words, which can assist the model in learning better language structures and semantic information in sentences, thereby improving its performance [12].

This model can learn deep language expression capabilities through these two tasks, resulting in a more accurate text representation.

The BERT model's fine-tuning model involves applying the pre-trained generic text representation to specific Natural Language Processing tasks and adapting it through fine-tuning. To improve performance, the model's structure or parameters can be adjusted during this process based on the specific task requirements.

BERT is very suitable for sentiment analysis because it captures the context and meaning of words in sentences [54].

RoBERTa (Robustly Optimized BERT Approach) is a language model based on BERT model [55]. It has optimized and improved the BERT model by improving pre-training method, using large data, and computing capability [56]. RoBERTa model removes the next sentence prediction objective from BERT.

GPT (Generative Pre-trained Transformer) is a transformer-based model, which is first introduced by OpenAI in [57]. The training data for GPT-1 comes from a large amount of text on the internet, including Wikipedia, news articles, and novels.

The main distinction between GPT-2 and GPT-1 is size of the model and the scale of the training dataset. GPT-2 has more parameters and a larger model architecture than GPT-1, which improves the model's language generation capability. GPT-2 also employs some novel techniques to improve model performance, such as dynamic masking, unsupervised training, and multi-stage fine-tuning.

GPT-3 has gotten a lot of attention because its better performance in different NLP tasks, especially for its versatile in-context few-shot learning ability.

Using GPT-3 for sentiment analysis can improve accuracy by leveraging its natural language generation capabilities, automatically generating sentiment scores for new documents to achieve automated sentiment analysis, and fine-tuning on task-specific datasets using transfer learning. As a result, there is existing literature that employs GPT for performing sentiment analysis.

Google proposed a language generation model called T5 model [58]. The T5 model converts various tasks into a unified text-to-text format, allowing model to learn and predict using a universal text representation.

T5 model is trained in 2 phases: pre-training and fine-tuning. T5 model is trained on a large corpus of text data in the pre-training stage, basing on the Transformer's encoder and decoder architecture to learn a general text representation.

TABLE III. COMPARISON OF TRANSFORMER LANGUAGE MODELS

Models	**Parameter (Millions)**	**Dataset**	**Method**	**Accuracy (%)**
BERT	Base: 110 Large: 340	BooksCorpus wiki	BERT (with MLM and NSP)	93.6
RoBERTa	Base: 110 Large: 340	BooksCorpus Wiki CC-News Stories	BERT (without NSP)	94.2
GPT-2	1.5B	WebText	LM	66.0
GPT-3	175B	Common Crawl, WebText 2, Books1, Books2, Wikipedia	Caption prediction	79.3
T5	Base: 220 Large: 770	Colossal Clean Crawled Corpus (C4)	Text Infilling	94.6
BLOOM	176B	Multilingual dataset	LM	-

T5 model is fine-tuned on a specific task to achieve better performance during the fine-tuning phase.T5 model is be used in many research, usually compared with other models [56].

Bloom model refers to a large, multilingual language model developed by BigScience. It contains 176 billion parameters and was trained for 3.5 months [59].

III. RESEARCH METHODOLOGY AND SYSTEM ARCHITECTURE

A. *Research Methodology*

To ensure scientific rigor, the research method employed by this studied is based on development methodology (Nunamaker Jr et al., 1990). The methodology is divided into five major processes: (1) construct a conceptual framework, (2) develop a system architecture, (3) analyze and design the system, (4) build the system, and (5) observe and evaluate the system.

1. *Construct a Conceptual Framework*

   According to introduction in Chapter 1, the research motivation and problem have been identified. After conducting a literature review in Chapter 2, it is possible to analyze more clearly the application of AI in carbon market sentiment text, system functionalities and requirements, understand the system building process, and study the relevant criteria and methods for system development.

2. *Develop a System Architecture*

   Construct a modular and scalable system architecture for using artificial intelligence in carbon market sentiment indicator research. Define the system's core functionalities and describe their interrelationships.

3. *Analyze and Design the System*

Design a database can implement the functionalities of the carbon market sentiment text analysis system, and choose the best feasible solution from multiple options.

4. *Build the System*

In the process of building the carbon market sentiment indicator system, further understanding of the core, architecture, and design concepts, as well as insights into the problem. Enterprises can utilize text analysis to extract social media content related to carbon market sentiment. This enables them to quickly browse the text and understand the situation through accurate analysis results, uncovering relevant issues.

5. *Observe and Evaluate the System*

By conducting study to observe use of the carbon market sentiment indicator system, as well as using system simulation and experimental methods to evaluate the system's performance, it can help policymakers develop appropriate tax policies and improve people's understanding of environmental policies, thus further increasing trust in government, education, and views on the impact of taxes on individuals and businesses.

B. *Proposed System Architecture*

In our research, we will propose the Carbon Market Sentiment Index (CMSI) with a transformer-based model. The research architecture will be shown in Fig.1. There will be three major modules for constructing the CMSI.

First, we will collect data from news websites and social media platforms. As we construct the CMSI, we will analyze textual data from Twitter and other social media platforms. We will choose Google News as our primary news source due to its comprehensive coverage of news from various outlets. From each article, we will collect fields such as the headline, publication date, content, author, and source outlet. These fields will allow us to have a detailed context of the news, ensuring a more nuanced sentiment analysis. This textual data will provide insights not only into market dynamics and investor sentiments but also into governmental policies, industry news, and other influential factors on market sentiment. Subsequently, we will proceed with data preprocessing, which will include cleaning up invalid data, tokenization, removing stop words, among other steps, to ensure data quality and enhance analytical accuracy.

Second, we will utilize a transformer-based model to classify each document or review as positive, negative, or neutral. We will choose BERT for its deep understanding of textual context, making it ideal for sentiment analysis. GPT, as a decoder, will excel in generating sentiment-rich reviews. Combining these models will enhance our accuracy in analyzing market sentiment shifts. Afterward, we will generate a daily time series with the polarity scores.

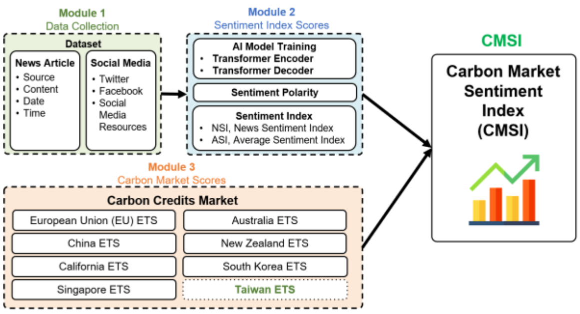

Fig. 1. The Proposed System Architecture of Carbon Market Sentiment Index (CMSI)

Third, we will refer to the European Union Emission Trading System, China's Carbon Emission Trading System, Singapore's Carbon Emission Trading System, and Taiwan Emission Trading System. We will then use two sentiment indexes to measure market participants' attitude toward the carbon market. Lastly, we will create a CMSI indicator to provide investors with a reference.

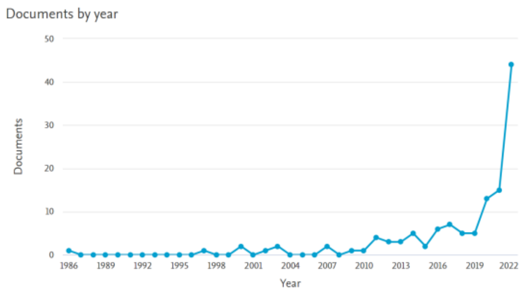

Fig. 2. The trend analysis on the research of carbon and sentiment analysis

IV. DATA ANALYSIS AND DISCUSSION

In this section, we proceed the trend analysis, keywords analysis, region analysis, and published journals analysis on carbon market sentiment research from Scopus for constructing the architecture of Carbon Market Sentiment Index (CMSI).

A. *Trend Analysis*

In the past few years, there has been a lot of studies on carbon and sentiment analysis. We searched for relevant literature from the database of Scopus. We use carbon, and sentiment as keyword to search. It shows a total of 83

TABLE IV. SEARCH COMMANDS FOR SCOPUS DATABASE

Step	Search Commands	Results
(1) keywords: carbon market + sentiment analysis	(TITLE-ABS-KEY(carbon AND market) AND TITLE-ABS-KEY(sentiment AND analysis))	18
(2) synonyms: carbon market/sustainable finance/emissions/carbon + sentiment analysis	(TITLE-ABS-KEY(carbon AND market) OR TITLE-ABS-KEY(sustainable AND finance) OR TITLE-ABS-KEY(carbon) OR TITLE-ABS-KEY(emissions) AND TITLE-ABS-KEY(sentiment AND analysis))	130
(3) scope time: 2019~2023	(TITLE-ABS-KEY(carbon AND market) OR TITLE-ABS-KEY(sustainable AND finance) OR TITLE-ABS-KEY(carbon) OR TITLE-ABS-KEY(emissions) AND TITLE-ABS-KEY(sentiment AND analysis)) AND (LIMIT-TO (PUBYEAR,2023) OR LIMIT-TO (PUBYEAR,2022) OR LIMIT-TO (PUBYEAR,2021) OR LIMIT-TO (PUBYEAR,2020) OR LIMIT-TO (PUBYEAR,2019))	96

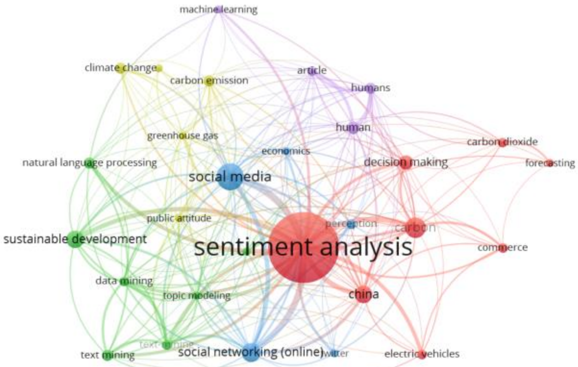

Fig. 3. Co-occurrence Keyword Visualization for Research

document results. Figure 2 illustrates the number of literatures has increased significantly since 2019. Many scholars have started to concentrate on sentiment analysis for research in carbon fields.

B. Keyword Analysis

We searched for relevant literature from the database of Scopus. These databases were chosen because they provide the greatest coverage.

In order to choose the most appropriate search keywords, the study discovered a pattern of keyword occurrence in the academic literature on purchase prediction by an initial scoping search of articles directly linked to the issue. Then, the study scanned the titles and keywords of the searched literature to find out the frequently used keywords. The search keywords were iteratively developed to enrich the literature search results and to omit the unrelated literature.

Linnenluecke [60] argued that the systematic literature review research should be replicable. Thus, we record the search codes as Table IV.

The current study searches the literature using the Boolean operator. Carbon and sentiment analysis are the keywords utilized in this study.

There were a total of 96 articles included. The Appendix contains a comprehensive listing of the articles included in the study. Figure 3 visualizes the co-occurrence keyword with carbon market and sentiment analysis.

C. Region Analysis

Table V shows the top ten country from the research affiliation in this research. The countries with the most literatures come from China, United States, and United Kingdom. This is why we have listed the Chinese and EU carbon markets as our research objects.

D. Published Research Source Analysis

Table VI lists the distribution of these documents, including the types of journals to which researchers most frequently submit and the corresponding number of manuscripts. We better understand the core of the topic related to this research in the academic field, as well as related literature information.

V. CONCLUSION

After systematic literature review of related knowledge, we introduced the concept of applying sentiment analysis technology to social media platforms. And reviewed its categories, tasks, and application. Then illustrate pre-trained language model, sustainable finance, and sentiment index. We designed our research proposal in chapter 3 and explained its process. In addition, we will use the score to construct a sentiment indicator. It will present in a time series to visualize the public market's concern about carbon issues.

TABLE V. documents per year by source

Rank	Country	Count	Percentage
1	China	33	34%
2	United States	13	14%
3	United Kingdom	9	9%
4	Germany	8	8%
5	India	5	5%
6	Canada	4	4%
7	Denmark	4	4%
8	Hong Kong	3	3%
9	Australia	2	2%
10	Bangladesh	2	2%

TABLE VI. top 10 affiiation country in this research

Rank	Source	Count
1	Sustainability Switzerland	9
2	Lecture Notes In Networks And Systems	4
3	Communications In Computer And Information Science	3
4	Energies	2
5	Environmental Impact Assessment Review	2
6	Frontiers In Environmental Science	2
7	International Review Of Financial Analysis	2
8	Lecture Notes In Electrical Engineering	2
9	Sustainable Production And Consumption	2
10	Technological Forecasting And Social Change	2
11	Applied Economics Letters	1
12	Applied Energy	1

CMSI is an advanced sentiment analysis tool that deeply analyzes financial news and social media, providing holistic insights into the carbon market's emotional dynamics. It assists financial institutions in identifying investment opportunities and risks, facilitating refined investment strategies. CMSI enables investors to intuitively understand market trends, promoting informed investment decisions. Our future research, enriched with real-world case studies, aims to vividly demonstrate CMSI's practical value and influence.

ACKNOWLEDGMENT

This research was supported in part by the National Science and Technology Council (NSTC), Taiwan, under grants NSTC 112-2425-H-305-002- and NSTC 112-2627-M-038-001-, and National Taipei University (NTPU), Taiwan under grants 112-NTPU-ORDA-F-003 and 112-NTPU-ORDA-F-004.

REFERENCES

[1] Y. Liu, J. Lu, J. Yang, and F. Mao, "Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax," *Mathematical Biosciences and Engineering,* vol. 17, no. 6, pp. 7819-7837, 2020.

[2] S. Peng *et al.*, "A survey on deep learning for textual emotion analysis in social networks," *Digital Communications and Networks,* vol. 8, no. 5, pp. 745-762, 2022.

[3] J. Xiao and X. Luo, "Aspect-level sentiment analysis based on BERT fusion multi-attention," in *2022 14th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2022: IEEE, pp. 32-35.

[4] K. Chatsiou, "Text classification of manifestos and COVID-19 press briefings using BERT and convolutional neural networks," *arXiv preprint arXiv:2010.10267*, 2020.

[5] Y. Zheng, M. Zhou, and F. Wen, "Asymmetric effects of oil shocks on carbon allowance price: evidence from China," *Energy Economics*, vol. 97, p. 105183, 2021.

[6] M. Edalati, A. S. Imran, Z. Kastrati, and S. M. Daudpota, "The potential of machine learning algorithms for sentiment classification of students' feedback on MOOC," in *Intelligent Systems and Applications: Proceedings of the 2021 Intelligent Systems Conference (IntelliSys) Volume 3*, 2022: Springer, pp. 11-22.

[7] J. Biswas, M. M. Rahman, A. A. Biswas, M. A. Z. Khan, A. Rajbongshi, and H. A. Niloy, "Sentiment analysis on user reaction for online food delivery services using bert model," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2021, vol. 1: IEEE, pp. 1019-1023.

[8] K. Zhou and Y. Li, "Carbon finance and carbon market in China: Progress and challenges," *Journal of Cleaner Production,* vol. 214, pp. 536-549, 2019.

[9] B. Su and J. Peng, "Sentiment Analysis of Comment Texts on Online Courses Based on Hierarchical Attention Mechanism," *Applied Sciences,* vol. 13, no. 7, p. 4204, 2023.

[10] H. W. Kang, K. K. Chye, Z. Y. Ong, and C. W. Tan, "The Science of Emotion: Malaysian Airlines Sentiment Analysis using BERT Approach," in *International Conference on Digital Transformation and Applications (ICDXA)*, 2021, vol. 25, p. 26.

[11] M. Liu and Q. Ying, "The role of online news sentiment in carbon price prediction of China's carbon markets," *Environmental Science and Pollution Research,* pp. 1-9, 2023.

[12] H. Li, Y. Ma, Z. Ma, and H. Zhu, "Weibo text sentiment analysis based on bert and deep learning," *Applied Sciences,* vol. 11, no. 22, p. 10774, 2021.

[13] O. Figueira, Y. Hatori, L. Liang, C. Chye, and Y. Liu, "Understanding COVID-19 public sentiment towards public health policies using social media data," in *2021 IEEE Global Humanitarian Technology Conference (GHTC)*, 2021: IEEE, pp. 8-15.

[14] F. Wen, H. Zhao, L. Zhao, and H. Yin, "What drive carbon price dynamics in China?," *International Review of Financial Analysis,* vol. 79, p. 101999, 2022.

[15] R. Cai *et al.*, "Sentiment analysis about investors and consumers in energy market based on BERT-BiLSTM," *IEEE access,* vol. 8, pp. 171408-171415, 2020.

[16] Y. Li and N. Li, "Sentiment analysis of Weibo comments based on graph neural network," *IEEE Access,* vol. 10, pp. 23497-23510, 2022.

[17] N. Xiang, L. Wang, S. Zhong, C. Zheng, B. Wang, and Q. Qu, "How does the world view China's carbon policy? A sentiment analysis on Twitter data," *Energies,* vol. 14, no. 22, p. 7782, 2021.

[18] I. Aygün, B. Kaya, and M. Kaya, "Aspect based twitter sentiment analysis on vaccination and vaccine types in covid-19 pandemic with deep learning," *IEEE Journal of Biomedical and Health Informatics,* vol. 26, no. 5, pp. 2360-2369, 2021.

[19] R. Chandra and A. Krishna, "COVID-19 sentiment analysis via deep learning during the rise of novel cases," *PloS one,* vol. 16, no. 8, p. e0255615, 2021.

[20] Y. Zhao, C. Wang, Y. Sun, and X. Liu, "Factors influencing companies' willingness to pay for carbon emissions: Emission trading schemes in China," *Energy Economics,* vol. 75, pp. 357-367, 2018.

[21] V. Kharde and P. Sonawane, "Sentiment analysis of twitter data: a survey of techniques," *arXiv preprint arXiv:1601.06971,* 2016.

[22] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electronics,* vol. 9, no. 3, p. 483, 2020.

[23] Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review," *Procedia Computer Science,* vol. 161, pp. 707-714, 2019.

[24] S. Kumar, D. Sharma, S. Rao, W. M. Lim, and S. K. Mangla, "Past, present, and future of sustainable finance: Insights from big data analytics through machine learning of scholarly research," *Annals of Operations Research*, pp. 1-44, 2022.

[25] S. P. Ferris and K. P. Rykaczewski, "Social investing and portfolio management," *Business & Society,* vol. 25, no. 1, pp. 1-7, 1986.

[26] M. Al Muhairi and H. Nobanee, "Sustainable financial management," *Available at SSRN 3472417*, 2019.

[27] M. Fleschutz, M. Bohlayer, M. Braun, G. Henze, and M. D. Murphy, "The effect of price-based demand response on carbon emissions in European electricity markets: The importance of adequate carbon prices," *Applied Energy,* vol. 295, p. 117040, 2021.

[28] J. Ye and M. Xue, "Influences of sentiment from news articles on EU carbon prices," *Energy Economics,* vol. 101, p. 105393, 2021.

[29] F. R. Aune and R. Golombek, "Are carbon prices redundant in the 2030 EU climate and energy policy package?," *The Energy Journal,* vol. 42, no. 3, 2021.

[30] A. D. Ellerman and P. L. Joskow, *The European Union's emissions trading system in perspective*. Pew Center on Global Climate Change Arlington, VA, 2008.

[31] P. Yu, Y. Weng, and A. Ahuja, "Carbon Financing and the Sustainable Development Mechanism: The Case of China," in *Handbook of Research on Energy and Environmental Finance 4.0*: IGI Global, 2022, pp. 301-332.

[32] C. Pan *et al.*, "Key challenges and approaches to addressing barriers in forest carbon offset projects," *Journal of Forestry Research,* vol. 33, no. 4, pp. 1109-1122, 2022.

[33] X. Liu and Z. Lin, "Impact of COVID-19 pandemic on electricity demand in the UK based on multivariate time series forecasting with bidirectional long short term memory," *Energy,* vol. 227, p. 120455, 2021.

[34] A. Giachanou and F. Crestani, "Like it or not: A survey of twitter sentiment analysis methods," *ACM Computing Surveys (CSUR),* vol. 49, no. 2, pp. 1-41, 2016.

[35] R. Wagh and P. Punde, "Survey on sentiment analysis using twitter dataset," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2018: IEEE, pp. 208-211.

[36] K. Ravi, V. Siddeshwar, V. Ravi, and L. Mohan, "Sentiment analysis applied to Educational Sector," in *2015 IEEE international conference on computational intelligence and computing research (ICCIC)*, 2015: IEEE, pp. 1-6.

[37] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review,* vol. 55, no. 7, pp. 5731-5780, 2022.

[38] V. Bonta and N. K. a. N. Janardhan, "A comprehensive study on lexicon based approaches for sentiment analysis," *Asian Journal of Computer Science and Technology,* vol. 8, no. S2, pp. 1-6, 2019.

[39] A. Sadia, F. Khan, and F. Bashir, "An overview of lexicon-based approach for sentiment analysis," in *2018 3rd International Electrical Engineering Conference (IEEC 2018)*, 2018, pp. 1-6.

[40] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Systems,* vol. 226, p. 107134, 2021.

[41] K. Elshakankery and M. F. Ahmed, "HILATSA: A hybrid Incremental learning approach for Arabic tweets sentiment analysis," *Egyptian Informatics Journal,* vol. 20, no. 3, pp. 163-171, 2019.

[42] J. Bernabé-Moreno, A. Tejeda-Lorente, J. Herce-Zelaya, C. Porcel, and E. Herrera-Viedma, "A context-aware embeddings supported method to extract a fuzzy sentiment polarity dictionary," *Knowledge-Based Systems,* vol. 190, p. 105236, 2020.

[43] A. I. M. Elfeky, T. S. Y. Masadeh, and M. Y. H. Elbyaly, "Advance organizers in flipped classroom via e-learning management system and the promotion of integrated science process skills," *Thinking Skills and Creativity,* vol. 35, p. 100622, 2020.

[44] V. Kuleto *et al.*, "Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions," *Sustainability,* vol. 13, no. 18, p. 10424, 2021.

[45] J. Zhou and J.-m. Ye, "Sentiment analysis in education research: a review of journal publications," *Interactive learning environments,* pp. 1-13, 2020.

[46] Q. A. Xu, V. Chang, and C. Jayne, "A systematic review of social media-based sentiment analysis: Emerging trends and challenges," *Decision Analytics Journal,* p. 100073, 2022.

[47] Z. Kastrati, L. Ahmedi, A. Kurti, F. Kadriu, D. Murtezaj, and F. Gashi, "A deep learning sentiment analyser for social media comments in low-resource languages," *Electronics,* vol. 10, no. 10, p. 1133, 2021.

[48] D. Araci, "Finbert: Financial sentiment analysis with pre-trained language models," *arXiv preprint arXiv:1908.10063,* 2019.

[49] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems,* vol. 30, 2017.

[50] L. Mathew and V. Bindu, "A review of natural language processing techniques for sentiment analysis using pre-trained models," in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, 2020: IEEE, pp. 340-345.

[51] S. Arroni, Y. Galán, X. Guzmán-Guzmán, E. R. Núñez-Valdez, and A. Gómez, "Sentiment Analysis and Classification of Hotel Opinions in Twitter With the Transformer Architecture," 2023.

[52] M. G. Sousa, K. Sakiyama, L. de Souza Rodrigues, P. H. Moraes, E. R. Fernandes, and E. T. Matsubara, "BERT for stock market sentiment analysis," in *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, 2019: IEEE, pp. 1597-1601.

[53] M. D. Deepa, "Bidirectional encoder representations from transformers (BERT) language model for sentiment analysis task," *Turkish Journal of Computer and Mathematics Education (TURCOMAT),* vol. 12, no. 7, pp. 1708-1721, 2021.

[54] S. Alaparthi and M. Mishra, "BERT: A sentiment analysis odyssey," *Journal of Marketing Analytics,* vol. 9, no. 2, pp. 118-126, 2021.

[55] Y. Liu *et al.*, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[56] K. Pipalia, R. Bhadja, and M. Shukla, "Comparative analysis of different transformer based architectures used in sentiment analysis," in *2020 9th International Conference System Modeling and Advancement in Research Trends (SMART)*, 2020: IEEE, pp. 411-415.

[57] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018.

[58] C. Raffel *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *The Journal of Machine Learning Research,* vol. 21, no. 1, pp. 5485-5551, 2020.

[59] T. L. Scao *et al.*, "Bloom: A 176b-parameter open-access multilingual language model," *arXiv preprint arXiv:2211.05100,* 2022.

[60] M. K. Linnenluecke, "Resilience in business and management research: A review of influential publications and a research agenda," *International Journal of Management Reviews,* vol. 19, no. 1, pp. 4-30, 2017.