**ANL 488 PROJECT PROPOSAL**

**CHINA'S GREEN FINANCE INSIGHTS FROM CHINESE LITERATURE: A TEXT MINING APPROACH**

****

**Submitted by**

LIM WEN QI

E2081571

**SCHOOL OF BUSINESS**

**Singapore University of Social Sciences**

**Presented to Singapore University of Social Sciences in partial fulfilment of the requirements for the Degree of Bachelor of Science in Business Analytics**

2023

**TABLE OF CONTENTS**

Chapter One Introduction 1

Chapter Two Literature Review 5

Chapter Three Data Understanding and Preparation 10

Chapter Four Proposed Modelling and Evaluation 18

Chapter Five Proposed Schedule 20

References 21

**INTRODUCTION**

China’s rapid economic and societal development in recent years have resulted in large amounts of environmental pollution and ecological problems. Accounting for 27% of global carbon dioxide and one-third of greenhouse gas emissions (The World Bank, 2022), China plays a defining role in its contribution towards climate change. This concern was emphasised during the 2015 Paris Agreement and 2016 United Nations Sustainable Development Goals, which increased China’s awareness and motivation towards achieving sustainable development (Sun et al., 2023). Ever since, China has pledged carbon neutrality by 2060, with emissions peaking in 2030.

China identified that the main source of emissions was due to its outdated financial system that relied heavily on fossil fuels for economic development. To tackle this problem, China’s leadership has emphasised the importance of green finance for its future economic development (Jiang et al., 2018). The nation defined green finance as support for environmentally beneficial projects, green financial products and financial services that support the financing of green projects (Peng et al., 2018).

In 2016, China established a novel green financing system inclusive of green bonds, credit and development funds. The system involved green finance regulations issued at the provincial level and green innovations conducted in designated pilot zones (PBC et al., 2016). This marked the start of China becoming the fastest-growing green finance market in the world. China’s green bond market grew from 0% before 2015 to the largest in the world in 2016 (Jiang et al., 2018). In 2022, the nation issued the highest amount of green bonds globally, totalling US$76.25 billion (Wu & Ahmad, 2023). At the end of quarter 1 this year, green loans in domestic and foreign currencies exceeded 25 trillion yuan (US$3.5 trillion) and green bond balance surpassed 1.5 trillion yuan (US$205 billion) (The State Council, 2023a). As stated by Yi Gang (2023), the governor of the state bank People’s Bank of China, this figure “ranks among the top of the world”. This trend is expected to continue, with domestic commercial green bonds issued in the second quarter of this year surging by 88.26% year on year (The State Council, 2023b).

Despite its late start as compared to countries in the European region and United States (Tang, 2021), China managed to quickly and successfully develop its green finance market. This indicates that there are many aspects to be learnt from the nation’s success. In order to understand China’s strategy more effectively, an analysis of nation-related green finance materials can be conducted. Existing bibliometric analysis had looked into China’s green policies (Liao & Chen, 2020; Dikau & Volz, 2021) and green financial instruments (Feng et al., 2023; Larsen, 2023). However, a limitation is that most studies only reviewed materials related to China that are published in English. With the first language used in China being Chinese, there is a high possibility that not all journal articles published in the country are made available in English literature databases (Shu et al., 2019). Even if research was conducted in Chinese databases (Li et al., 2022; Li et al., 2023), search terms in Chinese databases were in English. According to East View Information Services, a leading provider of foreign databases, English queries in Chinese databases are not considered comprehensive (2023). In addition, the meaning of words in Chinese literature may be lost during translation (Jin, 2019). To address the literature gap, this study will aim to conduct a bibliometric analysis using machine learning methods for retrieved data on Chinese green finance literature. The study seeks to understand more about the following research queries:

RQ1: How recently has research begun into green finance in China?

RQ2: What are the overall green finance insights gathered from Chinese literature?

RQ3: What are the trending topics in green finance from Chinese literature?

This study is expected to provide substantial contributions to the current literature body. Firstly, the study will look into establishing when China researchers started researching about green finance. With the understanding that China only started its green finance market after 2015, it is intriguing to figure out if China had started research into the topic much earlier. Secondly, the study seeks to compile the main Chinese keywords obtained from co-occurrence analysis of all relevant literary titles obtained. To prevent loss in translation, findings will be presented in both Simplified Chinese and English. Thirdly, current trending topics from the past 3 years will be identified. The findings will serve as a reference point for international researchers to understand research trends in Chinese literature.

The study will involve academic article data retrieved from the China National Knowledge Infrastructure (CNKI) literature database. The CNKI database was chosen due to its reputation as the largest repository in China, collecting over 90% of resources that reflect latest developments in China’s economics, humanity, political social sciences, science and technology (East View Information Services, 2023). As the most authoritative and influential database in China, it is suitable for comprehensive retrieval of China’s literary content. To prevent non-comprehensiveness and loss in translation, the search queries for article retrieval will be entered in Simplified Chinese.

The methodology consists of a bibliometric analysis approach associated with machine learning methods. Analysis will be done using Natural Language Processing (NLP), a machine learning algorithm frequently used in text mining. NLP involves giving computers the ability to interpret and analyse language used by humans during communication (Goodfellow et al., 2016; Khurana et al., 2023). As an artificial intelligence model, NLP enables computers to effectively analyse large textual datasets. This shows the suitability of NLP as an algorithm for this study.

**LITERATURE REVIEW**

Green and sustainable finance has been gaining significant attention in the literary space globally over the last decade. This indicates that large amounts of literary articles have been present in the literature space.

Zhang et al. (2019) analysed 381 academic journals adopted from the Web of Science (WoS) core collection from 2001 to 2018. The journals were retrieved from a keyword search in WoS about green finance, climate finance and carbon finance. A bibliometric analysis was conducted using CiteSpace, a pattern visualisation software developed by Chen (2006) for previous green finance research. A clear upward trend of publications was observed during the sample period, with a notable increase since 2015. The field was noted to be dominated by science, climate change and environmental journals; with climate finance, climate change and policy identified as the top concerns. It was also noted that most works were published by authors from the United States, Europe and other developed countries. Despite being an emerging economy, it was noted that China authors ranked fourth for their contributions towards literature. Future research recommendations raised by the authors indicated that a further comprehensive review of relevant green finance literature is needed.

Maria et al. (2023) analysed 3275 documents from the Scopus database from 1990 to 2020. These documents included peer-reviewed articles, books, article reviews, and chapters that were queried using the keywords green finance, climate finance, carbon finance and sustainable finance. The growth rate of the field was 22%, with publications doubling every 3.4 years. Using a novel analytical method that comprises of complex network analysis, structural topic modelling (STM) and topological measures, Maria et al. discovered that literature was mainly segmented into 3 groups, with the most mature being about climate finance from developed to developing nations. Recent trends were found to be linked to climate financial risks, green bonds, and financial energy-emissions-economic models. Similar to Zhang et al., the authors noted that their study’s main limitation is the focus on the most influential articles in the field and proposed for further exploration of research on less meaningful and marginalised topics.

The above articles reviewed were based on a rather global perspective instead of being country-specific, providing insights on global literature trends. The following articles will provide insights specific to China.

Li et al. (2022) examined the logic of China’s green finance policy and proposed innovative research theories and methods. The study conducted analysis of articles obtained from China’s domestic database, CNKI, and international database, WoS. 1729 articles were retrieved from CNKI using the search term green finance. The source category was also set to include only articles from Peking University Core and CSSCI. This leaves areas for research into other sources. 1655 articles were retrieved from WoS using search terms green finance, environment finance, sustainable finance and green banking. After screening, 792 articles were retained.

Similar to Zhang et al. (2019), the study uses CiteSpace in their analysis. Domestic articles from CNKI noted that green finance, green credit and green bonds were the top three most popular research areas. Impact, performance and investment were found to be the three dominant areas in WoS articles. This indicates that there are different focus areas between domestic and international researchers about green finance in China. Domestic China researchers focussed mainly on green financial products and green transformation strategies. International researchers focussed mainly on climate change, corporate performance and investments. This emphasises the importance of analysing domestic articles in achieving a holistic understanding of green finance in China.

Feng et al. (2023) identified potential research avenues by understanding concerns about China’s green finance system. The study analysed 248 journal articles retrieved from the WoS database between 2013 and 2023. Authors of the study selected articles based on titles, author names, journals that the articles were published, author’s country of origin, and themes of interest with relevance to keywords such as green finance, impact, economic growth, China and CO2 emissions. Although a small dataset, the articles were ensured to be of high quality. Through cross checking of article titles and abstracts was conducted to exclude irrelevant articles and ensure the accuracy of findings gathered.

Similar to Zhang et al. (2019) and Li et al. (2022), CiteSpace was used during their analysis. Notable analysis consisted of keyword trends (1) and popular subject categories for citations (2). The findings are as summarised below:

1. Main topics and themes of research include green credit, green bonds and green investment. It was also interesting to discover that climate change, carbon emissions, and sustainable development were some of the top keywords. These findings indicate that China’s green finance research is multidisciplinary.
2. The dominant subject categories with the most citations were Environmental Science, Green and Sustainable Science and Technology, Computer Science, and Interdisciplinary Applications. It is insightful to understand that a financial topic like green finance is researched by authors mainly in the science and ecology fields. This re-emphasises that research into green finance is multidisciplinary.

The authors accessed that China’s green finance system has met with success, with consistent and substantial practical implementations. For even greater success, it was recommended that China’s government could establish a unified green finance framework with mandatory financial disclosures and promotion of green consumption incentives. Results from this study will investigate if these proposed recommendations are currently acted upon.

Li et al. (2022) and Feng et al. (2023) have provided multiple insightful findings that will guide my present study. However, both articles analysed their datasets using the CiteSpace software. CiteSpace has some limitations, with the possibility of cluster failure, inability to create heatmaps (Bujas et al., 2023), and the difficulty of interpreting visualisations (Synnestvedt et al., 2005). As such, my study will involve text mining using NLP coding techniques. Liao and Chen’s (2020) article uses various techniques that my study will implement.

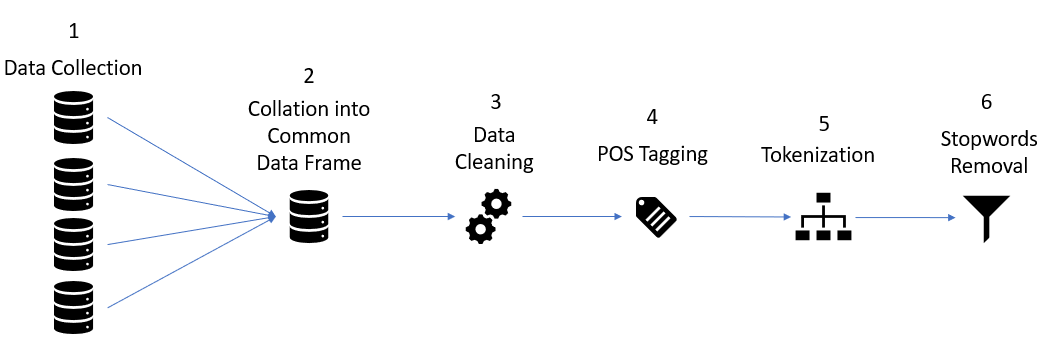
Liao and Chen (2020) seeked to understand China’s green finance industries and their developments over the period of 2017 to 2019. Using web scraping tool Bazhuayu on the Lawster website, the authors mined 1500 relevant green finance policy documents. The study used a text mining approach that consisted of forming a co-occurrence matrix and social semantic network to analyse China policy documents.

To prepare the dataset, stopwords were removed and a custom dictionary was introduced. This was because the standard stopwords and dictionary library was not able to provide enough context to the current research topic. The term frequency- inverse document frequency (TF-IDF) weighing system was adopted to form a co-occurrence matrix for popular keywords. TF-IDF measures the importance of a word or phrase by calculating its occurrence in a single document or a collection full of documents. This allowed the formation of a co-occurrence matrix, several word clouds and social semantic networks. The main keyword consistent in the majority of documents was innovation. This is in line with literature by PBC et al. (2016) discussed earlier in the introduction. Due to varying results from documents obtained in different years, it can be observed that the government’s focus areas different yearly. The authors noted that the policies vary over the years, with policies in 2017 and 2019 focusing on guidance; and policies in 2018 consolidating those published in 2017 for preparation of advancement in 2019.

As identified in the literature reviewed, there is a need for comprehensive retrieval of articles regarding the topic of green finance and analysis of China’s domestic databases using Chinese. This study will attempt to diminish these research gaps using Chinese NLP techniques.

**DATA UNDERSTANDING AND PREPARATION**

Figure 1 shows the six main steps that were used to obtain and prepare the dataset.



*Figure 1. Data preparation steps. Source: author’s construction*

Compared to English NLP, stemming and lemmatization are not required for Chinese NLP. Stemming and lemmatization involves reducing words to their base forms, also known as “lemmas”. However, Chinese phrases do not have a concept of stemming. As such, this step was not necessary.

Data preparation was conducted using Microsoft Office Excel in step 1 and Python from step 2 to step 6. The Python libraries used were namely pandas (Mckinney, 2011) and jieba.

*Step 1: Data Collection*

The CNKI database was scraped for academic articles with the search queries "可持续经济" (sustainable finance) and "绿色金融" (green finance). Both search queries generated search results in the areas of “科技” (science and technology) and “社科” (humanities and social sciences). Table 1 generalises the article count gathered.

*Table 1. Number of articles gathered from CNKI’s search engine*

|  |  |  |
| --- | --- | --- |
| Fields | Search Queries | |
| "可持续经济"  (Sustainable Finance) | "绿色金融"  (Green Finance) |
| “科技”  (Science and Technology) | 3,678 articles | 4,374 articles |
| “社科”  (Humanities and Social Sciences) | 9,879 articles | 10,027 articles |

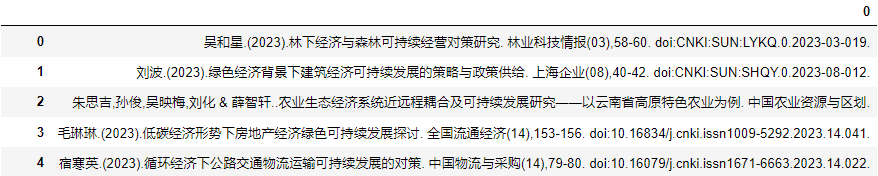
"可持续经济" (sustainable finance) retrieved 3,678 articles in the field of “科技” (science and technology) and 9,879 articles in the field of “社科” (humanities and social sciences). "绿色金融" (green finance) retrieved 4,374 articles in the field of “科技” (science and technology) and 10,027 articles in the field of “社科” (humanities and social sciences).

In order to obtain author, year, title and journal data from all queried articles, citations of all articles queried were obtained using CNKI’s inbuilt citation generator. Citations were gathered in the American Psychological Association (APA) format and collated into excel files. Each excel file contained data from different search queries and fields, amounting to four excel files. A web scraper tool was initially coded and deployed to mine the data queried from the database. However, the website had frequent timeouts and multiple captchas at random intervals. As a result, this idea was dismissed and data was gathered manually.

*Step 2: Data Collation*

All excel files were imported as dataframes into a jupyter notebook in a new Python 3 kernel. These processes were conducted by importing the pandas library and using the “read\_excel” method, with the arguments being each file’s path and headers set as None. Subsequently, dataframes that were retrieved using the same search query were concatenated to form a single file. This process was conducted using the pandas library method, concat. This resulted in two dataframes containing articles obtained from different search queries: "可持续经济" (sustainable finance) and "绿色金融" (green finance). Both dataframes were then concatenated using the same code as earlier mentioned, resulting in a single dataframe containing all articles from the search queries of interest.

As both "可持续经济" (sustainable finance) and "绿色金融" (green finance) were rather similar terminology, it is likely that the queries resulted in similar articles being queried. To prevent overlapping of articles in the dataframe, the duplicates inside the dataframe were dropped using the “drop\_duplicates” method located in the pandas library. A snippet of the resulting dataframe is shown in Figure 2.



*Figure 2. Snippet of common dataframe containing citations of all relevant articles*

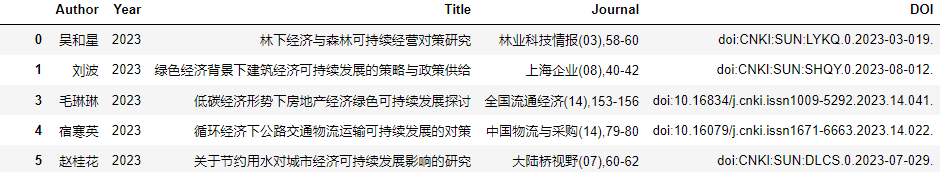
*Step 3: Data Cleaning*

The objective is to separate all data of interest into different dataframe columns. To do so, the initial idea was to split all rows in the dataframe by the delimiter “.”. However, some authors abbreviate their initials and multiple authors are indicated with “.”. Some examples include “R.加诺特” and “蔡群起,梁培金,林雯茹,苏锦荣,吴秀萍,吴婉羚... & 潘伟雄”. As such, a targeted approach was carried out.

The author and year were isolated from the title, journal and doi. This was done by using pandas’s “.str.split” method on the dataframe column header, 0, to get the first split using “)” . The arguments passed were pat = “)” and n=1. In the event this split resulted in any column being empty, this indicates that there is no year present in the article’s citation. For analysis of year during the modelling process, these rows were dropped using pandas’s “.dropna” method. Similar code using the “.str.split” method was used until all data was cleanly separated, with changes made to both the pat and n arguments. As there were multiple dataframes from the different variations of splitting the data, columns were added into a new dataframe and were assigned appropriate headers. This code is written as follows:

“cleaned\_dataset = cleaned\_dataset.assign(Author= split\_df\_author[0], Year= split\_df\_author[1], Title= split\_title\_journal\_doi[1], Journal= split\_title\_journal\_doi[2], DOI= split\_title\_journal\_doi[3]”

The “cleaned\_dataset” variable was first assigned to an empty dataframe. Columns of interest from multiple split dataframes were then assigned with appropriate names into the “cleaned\_dataset” dataframe, using the pandas “. assign” method. Figure 3 shows a snippet of the final cleaned dataframe.

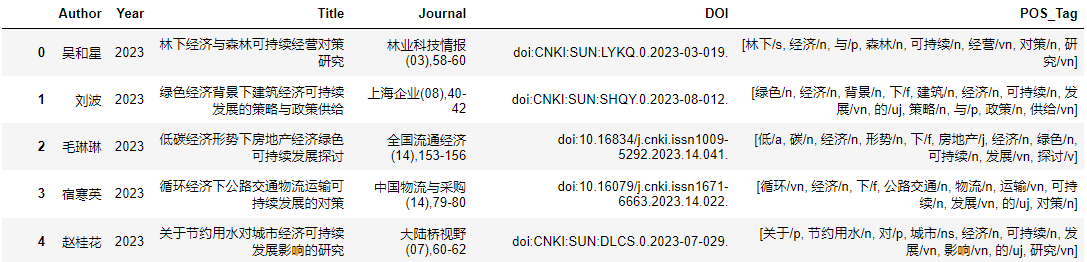


*Figure 3. Snippet of dataframe split by data columns of interest*

*Step 4: POS Tagging*

Part of speech (POS) Tagging was conducted to understand the context of words in sentences during text mining. It allows algorithms to better understand the grammatical structure of identified words. This process was conducted in step 4 before the removal of stopwords to ensure accurate understanding of context for the article titles.

The jieba library was imported into the jupyter notebook. Jieba is a Chinese text segmentation library that supports text-preprocessing methods such as POS tagging and word tokenization. Segmented words with their tags were stored as pairs in a list using the “jieba.posseg.lcut” method. The “use\_paddle” argument refers to an updated version of the tagger. POS tags were then identified by passing the “.apply” method across the dataframe “Title” column, with a lambda function passed with the code “jieba.posseg.lcut(x, use\_paddle = True)”. Figure 4 shows a snippet of the segmented title with its POS tags.



*Figure 4. Snippet of segmented title with POS Tags*

The POS tags follow rather closely with English terminology. The meanings of all tags from the original table in Chinese are reproduced in Table 2. Table 3 attempts to translate the tag meanings, and was formulated with close reference to Xing (2022).

*Table 2. Original POS Tags Meanings published by Jieba*

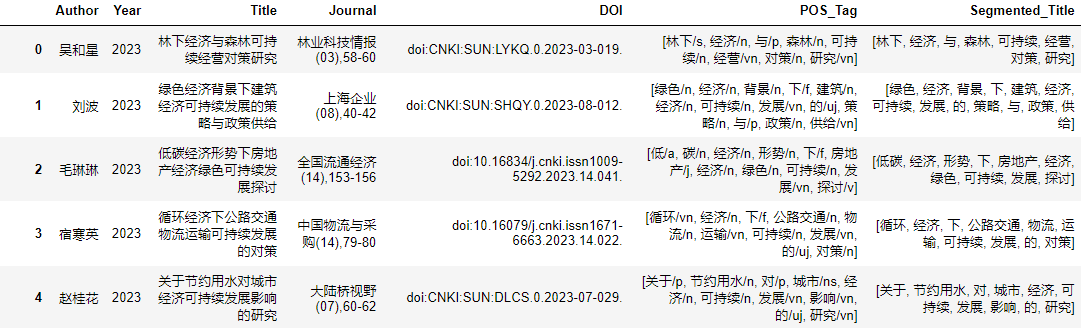
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 标签 | 含义 | 标签 | 含义 | 标签 | 含义 | 标签 | 含义 |
| n | 普通名词 | f | 方位名词 | s | 处所名词 | t | 时间 |
| nr | 人名 | ns | 地名 | nt | 机构名 | nw | 作品名 |
| nz | 其他专名 | v | 普通动词 | vd | 动副词 | vn | 名动词 |
| a | 形容词 | ad | 副形词 | an | 名形词 | d | 副词 |
| m | 数量词 | q | 量词 | r | 代词 | p | 介词 |
| c | 连词 | u | 助词 | xc | 其他虚词 | w | 标点符号 |
| PER | 人名 | LOC | 地名 | ORG | 机构名 | TIME | 时间 |

*Table 3. Translated POS Tags meanings used by Jieba*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Abbr. | Meaning | Abbr. | Meaning | Abbr. | Meaning | Abbr. | Meaning |
| n | Noun | f | Location noun | s | Location | t | Time |
| nr | Designation | ns | Name of place | nt | Name of Organisation | nw | Author’s name |
| nz | Other noun | v | Verb | vd | Verb adverb | vn | Verb noun |
| a | Adjective | ad | Adverb | an | Adjective noun | d | Adverb |
| m | Measurement | q | Quantifier | r | Pronoun | p | Preposition |
| c | Conjunction | u | Auxiliary verb | xc | Other function words | w | Punctuation |
| PER | Name | LOC | Name of Place | ORG | Name of organisation | TIME | Time |

*Step 5: Tokenization*

To get tokenized phrases from the data, the “.lcut” method from the jieba library was used. The argument of “cut\_all” can be used to specify the tokenisation mode, with “cut\_all = False” resulting in a more accurate output. The “.apply” method from pandas was applied to the dataframe’s “Title” column, with a lambda function passed with the code “jieba.lcut(x,cut\_all = False)”. Figure 5 shows a snippet of the segmented title.



*Figure 5. Snippet of cleaned segmented title with stopword removal*

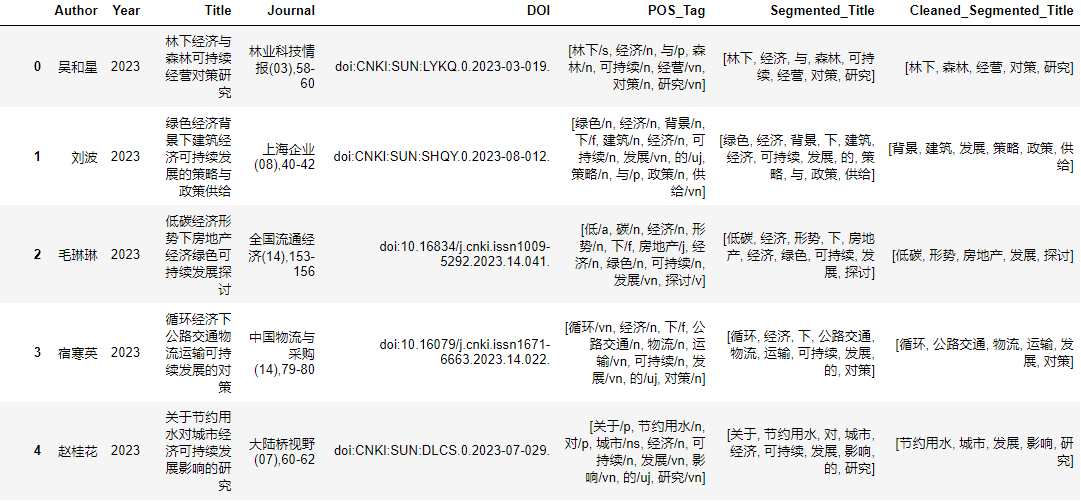
*Step 6: Stopwords Removal*

A text file containing stopwords was first created. Chinese stop words were obtained from the stopwordsiso github page and transferred into a text file. Variations of both search queries were also included as they will be frequent terms found in most titles. These include: 可持续经济 (sustainable finance), 可持续 (sustainable), 经济 (finance), 绿色金融 (green finance), 绿色 (green), 金融 (finance). Using the “open” build-in function, the text file was read into the notebook with its contents split by line and placed into a list named “stopwords”.

Using the “.apply” method from pandas, a lambda function that uses list comprehension was passed to include words that were not inside the “stopwords” list. These words were passed into a new column named "Cleaned\_Segmented\_Title". The code for this is as follows:

“cleaned\_dataset[“Cleaned\_Segmented\_Title”]= cleaned\_dataset[“Segmented\_Title”].apply(lambda lst: w for w in lst if w not in stopwords])”

Figure 6 shows a snippet of the cleaned segmented title.



*Figure 6. Snippet of cleaned segmented title with stopword removal*

**PROPOSED MODELLING AND EVALUATION**

Text mining has been gaining traction due to researchers analysing data from various fields for theory building (Kumar et al., 2021). This has led to an increase in the number of topic modelling techniques available. Upon comparing between various techniques such as the Latent Dirichlet Allocation (LDA), Non-negative Matrix Factorization (NMF), Top2Vec and BERTopic, BERTopic was found to be the most suitable for this study due to its exceptional ability of generating insights from short and unstructured text compared to other techniques (Egger & Yu, 2022).

BERTopic is a topic modelling technique that uses huggingface transformers and c-TF-IDF to create dense clusters for ease of topic interpretability. It uses a novel method of embedding representation, dimensionality reduction and extraction of topic representations using a custom class-based version of TF-IDF (Grootendorst, 2022) It is interesting to note that BERTopic encourages users to build their own topic model, with the algorithm being modular and maintains topic generation throughout a variety of sub-models. This makes it convenient for uses to adopt the algorithm at any major step of their choice.

With prior data pre-processing completed, the BERTopic model will be implemented from the topic representation step. Due to its independence from the dimensionality reduction step, the c-TF-IDF in this step can be freely altered. This can be carried out by explicitly passing “ctfidf\_model” into BERTopic to allow for parameter tuning and topic extraction customisation. There are two parameters of interest: “bm25\_weighting” and “reduce\_frequent\_words”. “bm25\_weighting = True” indicates that the class-based weighting method (log(1+A/fx)) has been changed to the BM-25 method (log(1+(A-fx + .5)/(fx+.5))), where fx indicates the frequency of word x across all classes and A indicates the average number of words per class. This is mainly used to increase robustness for small datasets towards stop words. “reduce\_frequent\_words = True” takes the square root of the term frequency after normalisation of the frequency matrix. This results in the removal of frequently occurring words that are not indicated as stopwords, but appear in every topic. For this study, the “bm25\_weighting” parameter will be set to False and “reduce\_frequent\_words” will be set as True.

After modelling is completed, visualisations will be generated. BERTopic provides multiple visualisations. For visualisation of topics, a method known as “.visualize\_topics” generates a plotly interactive graph for 2 dimensional representation. A topic similarity matrix can also be generated using the “.visualize\_heatmap” method. Term visualisation will also be conducted, with the “.visualize\_barchart” method showcasing the frequency of each term for each topic in bar graphs.

These modelling and visualisation steps seek to provide answers to RQ2 and RQ3, with RQ3 using a reduced dataset with articles from the recent 3 years.

The study will also adopt a different method when answering RQ1. Text mining will not be necessary to answer this question as the years are in numerical form. Instead, counts of articles from each year will be obtained and reported in a data table. A graph depicting article trends will also be plotted using the plotly library for better visualisation.

**PROPOSED SCHEDULE**

The proposed project timeline is as follows:

|  |  |
| --- | --- |
| Date | Tasks |
| 14-August-2023 to 8-September-2023 | Proposal Preparation |
| 8-September-2023 | Proposal Submission |
| 9-September-2023 to 11- September-2023 | Modelling process completed |
| 12- September-2023 to 15-September-2023 | Presentation slides completed |
| 16-September-2023 to 17-September-2023 | Practice before presentation |
| 18-September-2023 | Oral presentation of findings |
| 19-September-2023 to 15-October-2023 | Finish writing final report |
| 16-October-2023 to 29-October-2023 | Vetting of final report + Getting feedback |
| 30-October-2023 | Report Submission |

**REFERENCES**

Bujas, T., Vladimir, N., Koričan, M., Vukić, M., Ćatipović, I., & Fan, A. (2023). Extended Bibliometric Review of Technical Challenges in Mariculture Production and Research Hotspot Analysis. *Applied Sciences*, 13. https://doi.org/10.3390/app13116699

Chen, C. (2006, February 1). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J. Am. Soc. Inf. Sci*, 57(3) 359-377. https://doi.org/10.1002/asi.20317

Dikau, S., & Volz, U. (2021). Out of the window? Green monetary policy in China: window guidance and the promotion of sustainable lending and investment, *Climate Policy.* https://doi.org/10.1080/14693062.2021.2012122

East View Information Services. (2023). *China National Knowledge Infrastructure (CNKI) Frequently Asked Questions.* <https://www.eastview.com/resources/cnki-faq/#_Toc109472951>

Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology,* 7. https://doi.org/10.3389/fsoc.2022.886498

Feng, W., Bilivogui, P., Wu, J., & Mu, X. (2023). Green finance: Current status, development, and future course of actions in China. *Environmental Research Communications,* 5(3), 035005. <https://doi.org/10.1088/2515-7620/acc1c7>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure.* https://doi.org/10.48550/arXiv.2203.05794

Jiang, B., Guo, J., & Rhys, G. (2018, September). *Financial Services Special Report: China’s Green Finance Market.*

Jin, Z. (2019, January). The Limitation of Translatability Between English and Chinese. *Sino-US English Teaching*, 16(1) 12-23. https://doi.org/10.17265/1539-8072/2019.01.003

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications.* 82(3), 3713–3744. <https://doi.org/10.1007/s11042-022-13428-4>

Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights*, 1(1), 100008. https://doi.org/10.1016/j.jjimei.2021.100008

Larsen M. L. (2023). Bottom-up market-facilitation and top-down market-steering: comparing and conceptualizing green finance approaches in the EU and China. *Asia Europe journal*, 21(1), 61–80. https://doi.org/10.1007/s10308-023-00663-z

Li, J., Wen, K., Zhang, B., Dai, X., Qi, M., & Liu, B. (2023). Influence analysis and promotion countermeasures of green finance policy in China-Traceability based on grounded theory and validation using the csQCA method. *PloS one*, 18(5), e0285862. https://doi.org/10.1371/journal.pone.0285862

Li, J., Zhang, B., Dai, X., Qi, M., & Liu, B. (2022). Knowledge Ecology and Policy Governance of Green Finance in China-Evidence from 2469 Studies. *International journal of environmental research and public health*, 20(1), 202. https://doi.org/10.3390/ijerph20010202

Liao, Y., & Chen, J. (2020). Research on China’s Green Finance Policies Based on Text Mining. *E3S Web Conf*., 185, 02024. https://doi.org/10.1051/e3sconf/202018502024

Maria, M. R., Ballini, R., & Souza, R. F. (2023). Evolution of Green Finance: A Bibliometric Analysis through Complex Networks and Machine Learning. Sustainability, 15(2), 967. *MDPI AG*. http://dx.doi.org/10.3390/su15020967

Mckinney, W. (2011). pandas: a Foundational Python Library for Data Analysis and Statistics. *Python High Performance Science Computer.*

PBC, MOF, NDRC, MEP, CBRC, CSRC & CIRC (2016). *Guidelines for Establishing the Green Financial System. People’s Bank of China.* Ministry of Finance National Development and Reform Commission, Ministry of Environment Protection, China Banking Regulatory Commission, China Securities Regulatory, Commission China, Insurance Regulatory Commission.

Peng, H. , Luo, X., & Zhou, C. (2018) Introduction to China’s Green Finance System. *Journal of Service Science and Management,* 11, 94-100. https://doi.org/10.4236/jssm.2018.111009.

Shu, F., Julien, C. & Larivière, V. (2019). Does the web of science accurately represent chinese scientific performance?. *Journal of the Association for Information Science and Technology,* 70(10) 1138-1152. https://doi.org/10.1002/asi.24184

Sun, Y., Bao, Q., & Taghizadeh-Hesary, F. (2023). Green finance, renewable energy development, and climate change: evidence from regions of China. *Humanit Soc Sci Commun.,* 10(107). https://doi.org/10.1057/s41599-023-01595-0

Synnestvedt, M. B., Chen, C., & Holmes, J. H. (2005). CiteSpace II: visualization and knowledge discovery in bibliographic databases. AMIA ... Annual Symposium proceedings. *AMIA Symposium,* 724–728.

Tang, Y. J. (2021, April). Green And Social Finance Development Around the World. *Asian Development Outlook 2021: Financing a Green and Inclusive Recovery.*

The State Council. (2023a, June 9). *China's outstanding green loans reach 25 trillion yuan: PBOC governor.* The State Council The People's Republic of China. http://english.www.gov.cn/news/202306/09/content\_WS648290eac6d0868f4e8dcb84.html

The State Council. (2023b, June 23). *China's green finance development in the fast lane.* The State Council The People's Republic of China. http://english.www.gov.cn/news/202306/23/content\_WS649538d7c6d0868f4e8dd274.html

The World Bank. (2022, October 12). *China’s Transition to a Low-Carbon Economy and Climate Resilience Needs Shifts in Resources and Technologies.* https://www.worldbank.org/en/news/press-release/2022/10/12/china-s-transition-to-a-low-carbon-economy-and-climate-resilience-needs-shifts-in-resources-and-technologies#:~:text=China%20emits%2027%20percent%20of,of%20the%20world's%20greenhouse%20gases.

Wu, J., & Ahmad, R. (2023, February 12). *China to keep lead in green bond market amid alignment with global standards. A & P Global Market Intelligence.* <https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/china-to-keep-lead-in-green-bond-market-amid-alignment-with-global-standards-74039783#:~:text=China%20is%20expected%20to%20issue%20between%20%2490%20billion%20and%20%24100,and%20sustainability%20director%20at%20Deloitte>

Xing, W. (2022, Feburary 22). Parts of speech comparison table - part of speech tagging for natural language processing. https://www.cnblogs.com/bytesfly/p/part-of-speech.html

Yi Gang. (2023, June 22). Yi Gang: Proactively implementing the philosophy of green development to peak carbon emissions before 2030 and achieve carbon neutrality before 2060. *BIS.* https://www.bis.org/review/r230622h.htm#:~:text=At%20the%20end%20of%20Q1,the%20top%20of%20the%20world.

Zhang, D., Zhang, Z., & Managi, S. (2019). A bibliometric analysis on green finance: Current status, development, and future directions. *Finance Research Letters,* 29, 425–430. http://dx.doi.org/10.1016/j.frl.2019.02.003