Thesis_LI HONGLIN.pdf

by HONGLIN LI

Submission date: 27-Jun-2025 06:20AM (UTC-0700)
Submission ID: 2706863695
File name: Thesis_LI_HONGLIN.pdf (2.37M)
Word count: 10468
Character count: 58567

SENTIMENT ANALYSIS OF PUBLIC OPINION ON TRUMP'S 2025 CHINA TARIFF POLICY BASED ON "X"

LI HONGLIN

UNIVERSITI TEKNOLOGI MALAYSIA

PSZ 19:16 (Pind. 1/23)

UNIVERSITI TEKNOLOGI MALAYSIA
DECLARATION OF Choose an item.

Author's full name	:LI HONGLIN		
Student's Matric No.	:MCS241031	Academic Session	:202420252
Date of Birth	:30/12/1997	UTM Email	: lihonglin@graduate.utm.my
Choose an item. Title	:	SENTIMENT ANALYSIS OF PUBLIC OPINION ON TRUMP'S 2025 CHINA TARIFF POLICY BASED ON "X"	

I declare that this Choose an item. is classified as:

☒ **OPEN ACCESS** I agree that my report to be published as a hard copy or made available through online open access.

☐ **RESTRICTED** Contains restricted information as specified by the organization/institution where research was done. *(The library will block access for up to three (3) years)*

☐ **CONFIDENTIAL** Contains confidential information as specified in the Official Secret Act 1972)

(If none of the options are selected, the first option will be chosen by default)

I acknowledged the intellectual property in the Choose an item. belongs to Universiti Teknologi Malaysia, and I agree to allow this to be placed in the library under the following terms :

1. This is the property of Universiti Teknologi Malaysia
2. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of research only.
3. The Library of Universiti Teknologi Malaysia is allowed to make copies of this Choose an item. for academic exchange.

Signature of Student:

Signature :

Full Name
Date :

Approved by Supervisor(s)

Signature of Supervisor I:	Signature of Supervisor II
Full Name of Supervisor I	Full Name of Supervisor II
Date :	Date :

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

Date:

Librarian

Jabatan Perpustakaan UTM,

Universiti Teknologi Malaysia,

Johor Bahru, Johor

Sir,

CLASSIFICATION OF THESIS AS RESTRICTED/CONFIDENTIAL

TITLE: Click or tap here to enter text.

AUTHOR'S FULL NAME: Click or tap here to enter text.

Please be informed that the above-mentioned thesis titled _____ should be classified as RESTRICTED/CONFIDENTIAL for a period of three (3) years from the date of this letter. The reasons for this classification are

(i)

(ii)

(iii)

Thank you.

Yours sincerely,

SIGNATURE:

NAME:

ADDRESS OF SUPERVISOR:

"I hereby declare that I have read this proposal and in my opinion this proposal is sufficient in term of scope and quality for the award of the degree of Master of Data Science"

Signature : _____

Name of Supervisor I :

Date :

Signature : _____

Name of Supervisor II :

Date :

Signature : _____

Name of Supervisor III :

Date :

Declaration of Cooperation

This is to confirm that this research has been conducted through a collaboration LI HONGLIN and University Technologi Malaysia (UTM)

Certified by:

Signature :
Name :
Position :
Official Stamp
Date

* This section is to be filled up for theses with industrial collaboration

Pengesahan Peperiksaan

Tesis ini telah diperiksa dan diakui oleh:

Nama dan Alamat Pemeriksa Luar :

Nama dan Alamat Pemeriksa Dalam :

Nama Penyelia Lain (jika ada) :

Disahkan oleh Timbalan Pendaftar di Fakulti:

Tandatangan :

Nama :

Tarikh :

SENTIMENT ANALYSIS OF PUBLIC OPINION ON TRUMP'S 2025 CHINA TARIFF POLICY BASED ON "X"

LI HONGLIN

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Science (Data Science)

School of Computing
Faculty of Computing
Universiti Teknologi Malaysia

JUNE 2025

DECLARATION

I declare that this thesis entitled *" SENTIMENT ANALYSIS OF PUBLIC OPINION ON TRUMP'S 2025 CHINA TARIFF POLICY BASED ON "X" "* is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : .....

Name :

Date : 30 JUNE 2025

ACKNOWLEDGEMENT

In the process of completing this paper, I got a lot of people's care and help. Here, I would like to express my heartfelt thanks to all my friends who have given me guidance and support!

First of all, I would like to give my tutor Assoc. Professor. Dr. Mohd Shahizan bin Othman detailed guidance in the whole research process, especially valuable suggestions and patient explanations in methodology, so that I can carry out research work more systematically.

At the same time, I thank my predecessors for their help and guidance when I encountered technical problems. Your experience and suggestions pointed out the direction for me at the critical moment, which enabled me to advance all aspects of the research smoothly.

I am also very grateful to my friends and family. In the whole research process, it was your company and encouragement that taught me how to find a balance between study and life, reduce stress and keep the momentum of progress.

The successful completion of this paper is inseparable from the support and dedication of all the above people.

ABSTRACT

This study focuses on the public's emotional response to the Trump administration's tariff policy towards China in 2025 on social media "X" (formerly Twitter), and analyzes English tweets from January to May 2025 through NLP technology and VADER tools. Three-stage unsupervised analysis shows that negative emotions account for 60.4%, followed by positive emotions (28.4%) and neutral emotions (11.3%).

In order to optimize the classification accuracy, pseudo-labels are constructed based on VADER composite scores to form a balanced data set, and then machine learning models such as SVM and logistic regression are trained. Finally, the highest classification accuracy rate of 73.05% is achieved through the soft voting integrated model of SVM, logistic regression and gradient lifting.

A semi-supervised modeling path for sentiment analysis of policy-sensitive social media is proposed, which verifies the feasibility of pseudo-tagging without manual labeling. At the same time, it is found that the extreme fluctuation of trade policy is related to the polarization of public sentiment-the higher the policy intensity, the stronger the emotional response, and the space for rational discussion is compressed. The research results provide data and method reference for policy making, economic research and public opinion analysis. (Note: Students are allowed to use either single or one-and-a-half spacing for the abstract, as long as it fits within one page. The chosen spacing style must be consistent across both the English and Malay sections.)

Keywords: Sentiment analysis, Trump, tariffs on China, X platform, VADER, pseudo-label, Machine learning, Public opinion, semi-supervised learning

ABSTRAK

Kajian ini memfokuskan pada tindak balas emosi awam terhadap dasar tarif pentadbiran Trump ke atas China pada tahun 2025 di media sosial "X" (dahulunya Twitter), dan menganalisis ciputan bahasa Inggeris dari Januari hingga Mei 2025 menggunakan teknologi NLP dan alat VADER. Analisis tiga peringkat tanpa penyeliaan menunjukkan emosi negatif mendominasi sebanyak 60.4%, diikuti oleh emosi positif (28.4%) dan emosi neutral (11.3%).

Untuk mengoptimumkan ketepatan pengkelasan, label palsu dibina berdasarkan skor komposit VADER bagi membentuk set data seimbang, kemudian model pembelajaran mesin seperti SVM dan regresi logistik dilatih. Akhirnya, ketepatan pengkelasan tertinggi sebanyak 73.05% dicapai melalui model integrasi "soft voting" yang menggabungkan SVM, regresi logistik, dan "gradient boosting".

Satu laluan pemodelan separa berpenyelia bagi analisis sentimen media sosial sensitif dasar dicadangkan, yang membuktikan kebolehgunaan penandaan palsu tanpa label manual. Pada masa sama, didapati turun naik melampau dasar perdagangan berkaitan dengan polarisasi sentimen awam—semakin tinggi keamatan dasar, semakin kuat tindak balas emosi, dan ruang perbincangan rasional menjadi terhad. Hasil kajian memberikan rujukan data dan metodologi untuk pembentukan dasar, penyelidikan ekonomi, serta analisis pendapat awam.

Kata kunci: Analisis sentimen, Trump, tarif terhadap China, platform X, VADER, label palsu, pembelajaran mesin, pendapat awam, pembelajaran separa berpenyelia.

Nota: Pelajar dibenarkan menggunakan jarak baris tunggal atau satu setengah untuk abstrak, selagi muat dalam satu halaman. Gaya jarak baris yang dipilih perlu konsisten bagi bahagian bahasa Inggeris dan Melayu.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Introduction

In recent years, the Trump administration's capricious trade policies have caused widespread concern in global markets. Tariff is one of its important means. The move thus triggered retaliation from many countries in the face of additional tariffs imposed on China by the Trump administration in 2018 and new tariffs proposed in 2025 (Wengerek et al., 2025). It also sparked a heated debate on social media. Especially "X" (Twitter), which can provide real-time news and other information directly to the audience. In addition, the frequency of Trump's tweets on the social media platform "X" has gradually increased since he was first elected in November 2016. Among them, tweets containing "product" and "tariff" were found to have a negative impact on the stock market (Gjerstad et al., 2021). Twitter data has been used in a wide variety of applications, including disaster detection and localization (Loynes et al., 2022) and business trend forecasting.

Based on Twitter data, this study uses natural language processing and Sentiment analysis method to deeply explore the changes in public sentiment before and after the launch of Trump's tariff policy on China in 2025.

1.2 Problem Background

social media has played an increasingly important role in politics and economics, and Trump's tweets have received special attention. Previous research has shown that Trump's tweets about policy elicit different emotional responses from the public, for example, Dwianto et al. (2021) used automated sentiment analysis tools such as

Brand24 to find that Trump's policies implemented during the COVID-19 pandemic triggered more negative than positive sentiment.

In addition, it was found that analyzing the sentiment of Trump's tweets is predictive. Negative sentiment in tweets correlates with negative market reactions and can be used to infer market trends and public opinion (Pham et al., 2022). Sentiment brought about by social media has a sustained impact on the dynamics of international markets. Sentiment analysis has become an important means of monitoring market volatility during major events, as the resulting market volatility shows regular patterns of spikes (Abdollahi et al., 2024). Today, many studies mainly focus on the impact of Trump's tweets on finance and stock markets (Nishimura & Sun, 2025; Zhang et al., 2024).

However, there is still a gap in the research on public sentiment triggered by Trump's new 2025 tariffs. Therefore, this study will use X data to analyze and compare the static sentiment distribution before and after the introduction of the tariff policy, providing a new perspective for the study of this issue, as well as a new reference for policy making and economic development.

1.3 Problem Statement

It has been proved that tools and algorithms such as Support Vector Machine (SVM), Naive Bayes and VADER can well analyze whether the sentiment of social media text is positive or negative. For example, Faridzi et al. (2023) used an SVM combined with SMOTE algorithm to process Twitter data related to Indonesia's 2022 fuel price increase policy and were able to classify sentiment very accurately. Zangmo et al. (2024) analyzed the tweets of 2024 US presidential election using VADER model and classification algorithm and also found that naive Bayeas performs well in sentiment classification. Although there have been many sentiment analysis studies for major public events , there are still many gaps in current research on social media sentiment regarding Trump's new 2025 tariffs:

1. There is no analysis of the sentiment tendencies associated with this policy. There are no actual figures to show whether the public supports or opposes the policy.

2. It is uncertain how sentiment will change around the announcement. There is no data to support whether public sentiment changed from support to opposition or vice versa before and after the announcement of the policy.

Therefore, this study collects English tweets related to "Trump's tariffs on China 2025""on "X", performs text cleaning, uses tools such as VADER to determine whether the sentiment is positive or negative, calculates the proportion of different emotions, and then compares the differences in the distribution of emotions before and after the release of the policy, so as to provide data for constructing public opinion prediction models and evaluating policy responses in the future.

1.4 Research Aim and Objective

The study was designed to analyze public sentiment toward Trump's 2025 tariff policy on "X" to see whether they were predominantly for or against it, and to examine whether there was a significant change in public sentiment before and after the policy announcement.

The following steps will be conducted to achieve the objectives:

1. Collect English tweets related to "Trump tariffs" using an existing Twitter dataset.

2.Process the collected data, including cleaning text, converting formats, and filtering languages.

3.Use the VADER tool to determine whether the sentiment of the tweet is positive or negative.

4. VADER sentiment analysis is used to generate pseudo-labels and construct a semi-supervised sentiment classification model.

5. Count the number of different emotions and present them in a chart to see how people's emotions differ before and after the policy announcement.

1.5 Research scope

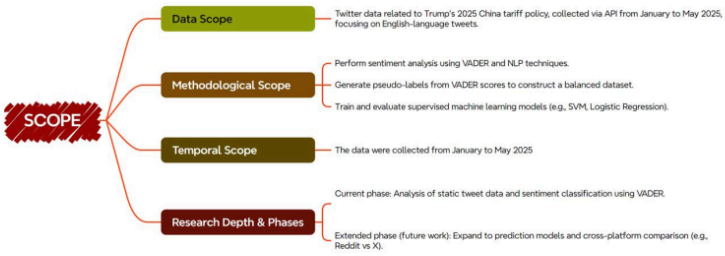
(a) Data scope: Trump tariff tweets before April 2025, mainly analyzing English tweets.

Since Twitter is tightening API access starting in 2023, we will initially use existing public data. Later, depending on the progress of the project, consider using simulated data or applying for advanced permissions to supplement the sample.

(b) Scope of methods: Mainly using open source pre-trained models such as Hugging Face and VADER on Kaggle.

A complete emotion recognition and classification process is established through parameter calling, label identification, result comparison and trend visualization.

(c) Research depth: Initial focus on static data analysis. Depending on the situation to extend to areas such as sentiment prediction compared to cross-platform. This paper mainly focuses on the description of emotional tendencies.



1.6 Contribution of the Project

1. Provide more practical cases of social media sentiment analysis to provide data for future academic research.

2. Create and display emotion maps so that decision makers can refer to them directly.

3. Provide data support for follow-up research, such as analyzing dynamic change trends or comparing different platforms

4. Based on VADER sentiment score, the pseudo-label method is used to construct a semi-supervised model, which enriches the sentiment classification method of social media data.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Tariff policy is one of the important means for government to regulate trade. For the country, it can protect industry and employment, adjust trade balance, and have an important impact on the public's price, employment, consumption choice and other aspects. This chapter provides a brief introduction to the content of Trump's 2025 tariffs on China and the impact of previous tariffs on China, as well as the application of sentiment analysis in social media, which provides a foundation for understanding the following chapters.

2.2 Content and impact of tariff policies on China

The tariff war between China and the US weakened the global trading system as early as the first term of the Trump administration. They cause huge economic losses in every country. Chinese imports fell 4.9 % and U.S. imports fell 4.5 %. Bilateral trade patterns are severely distorted (Zheng et al., 2023). Meanwhile, tariff announcements lead to negative (cumulative) average abnormal stock returns. (Wengerek et al., 2025), the uncertainty caused by trade tensions has a greater impact on the market than expected, and investor sentiment and information are also deeply affected by high uncertainty(Selmi et al., 2020). During this period, US tariffs on Chinese imports depressed Chinese demand for foreign inputs, adversely affecting third countries and causing an average GDP loss of 0.05% for the countries involved (Son, 2022).

In gey-Consumer Intelligence statistics show a spike in social media posts mentioning "tariffs" between January 1, 2024 and April 7, 2025. That's because the

Trump administration has enacted a series of tariffs since February 2025. In particular, since April, President Trump has announced the implementation of the so-called "reciprocal tariff" measures against US trading partners, announcing that the US would impose a 125% tariff on Chinese products, plus a 20% "fentanyl tax", bringing the total tariff to 145%. In response to the US move, China quickly and decisively announced a series of countermeasures - raising tariffs on imports from the US from 84 per cent to 125 per cent. It was the first country in the world to reject "reciprocal tariffs". Trump's "reciprocal tariff" policy has triggered turmoil in global stock markets, hit Sino-US economic and trade relations, and caused rising prices and signs of economic recession in the United States. On May 12, China-US relations turned a corner. The two sides issued a joint statement on the Geneva economic and trade talks, and the two sides mutually reduced tariffs: the comprehensive tariff rate of the US on Chinese goods was reduced from 145% to about 30%, the tariff rate of 24% was suspended (the base rate of 10% was maintained), and specific tariffs imposed in early April were cancelled. At the same time, China will take reciprocal measures to reduce its tariffs on the United States from 140 percent to 10-45 percent. Since tariff measures are closely related to the lives and investments of the public, the public usually communicates the discussion of tariff policies through social media. "X" is a platform for users around the world to discuss hot issues in real time, and the user group is wide, so it has triggered heated discussions again.

2.3 Sentiment analysis

Sentiment analysis is an important application field of natural language processing (NLP), which aims at automatically identifying the subjective emotional tendency in the text and judging whether the sentence is positive, negative or neutral. It can help companies, institutions or researchers understand the trend of public attitudes and opinions. Emotional analysis has three main technical methods:

Dictionary-based: It is simple and easy to understand, but its generalization ability is poor.

Machine learning: Training models on labeled data to extract features for classification requires manual feature engineering and depends on data quality.

Deep learning; Using neural network to automatically learn semantic features, end-to-end processing, but the calculation cost is high.

2.4 Application of existing technology

Existing model framework research covers various social media sentiment analysis applications, including elections, general datasets, public health topics, COVID-19 vaccine tweets, as well as improving analysis accuracy and interpretability. The performance of the model was evaluated by accuracy and other indicators using the methods of VADER, decision tree, KNN, Naive Bayes, SVM, LSTM, Bert, SentiWordNet, LIWC-22, ChatGPT 4.0 and word cloud visualization technology.

The following table provides an overview of the research on different sentiment analysis models in social media applications：

methods and models	application	Advantages	Disadvantages	Reference
VADER + Decision tree /KNN/Naive Bayes	Sentiment analysis of the US election, using supervised classification to test the correctness of the model	Easy to use, suitable for short text; The accuracy of Naive Bayes is 60.69%	The method is preliminary and the accuracy is low；	Zangmo et al. (2024)
BERT、LSTM、SVM	Multiple models are used for sentiment analysis on structured and unstructured data	BERT achieves the best performance with 86% accuracy and a high F1 score.	It requires a large amount of computational resources and has a high complexity	Elmassry et al. (2024)

		Applicable to various data distributions		
VADER, LIWC-22, TEXT2DATA, ChatGPT 4.0	Compare the effectiveness of various automatic tools in negative review recognition, especially for long texts To compare the effectiveness of various automatic tools in negative review identification, especially for long texts	LIWC-22 is more suitable for imbalanced data and VADER is suitable for long annotations. The overall agreement with manual annotation is good	ChatGPT 4.0 performed the worst on this task	Gandy et al. (2025)
SAVSA （SentiWordNet + VADER）	It improves the sentiment recognition accuracy of COVID-19 vaccine tweets and is suitable for imbalanced data	Multi-stage combination model to improve accuracy; Applicable to imbalanced data in social media	These methods are complex and depend on preprocessing and dictionary quality	Chockalingam & Thambusamy (2024)
VADER + Word	It is used to	Enhanced	It does not	Chavan et al.

Cloud Visualization	improve explainability and user understanding, and contributes to an intuitive understanding of text features	visualization for easy presentation of user intuition	improve the performance of the algorithm itself, but only increases interpretability	(2024)

Table 2.1 Application model

The VADER model was chosen for this project because it is not only suitable for analyzing short texts in social media such as Twitter, but also stable in imbalanced datasets dominated by negative comments and easy to integrate with visualization tools such as Word Cloud and DAX. Easy to generate intuitive reports (Gandy et al., 2025; Chavan et al., 2024)

2.5 Research Gap

Previous studies have discussed the impact of Trump's tweets on the market and public sentiment, especially tweets about China, which directly affect the stock market and increase its volatility and trading volume. Positive sentiment on microfairs leads to increased excess returns in China's manufacturing sector (Guo et al., 2021) The consumer goods industry also showed negative returns when Trump's tweets were negative about the pandemic (Pham et al., 2022), but there is no literature specifically analyzing the changes in public sentiment triggered by his 2025 tariffs. At present, there is a lack of comparative analysis of public opinion changes before and after the policy release based on X platform data. Existing sentiment analysis models, such as "VADER", have not been applied or validated in the context of this particular event. English tweets related to "Trump's new tariff policy on China in 2025" on the "X" platform were collected through the API. It focuses on the analysis of public sentiment before and after the April 10 tariff measures against China and after the May 12 reciprocal tariff reductions. "As such,

this study will fill the gap in the existing quantitative analysis of social media sentiment swings in the context of specific political events.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter is about how to analyze what people feel towards Trump's China 2025 tariff policy on "X". It discusses the entire process of data acquisition and cleaning to determine the feelings using the VADER analysis tool. There are three central concerns this study will focus on: people's expectation of the policy's effects, policy announcements, and policy updates. It will provide real evidence of a shift in public sentiments.

3.2 Research Framework

The research framework follows a standard data science project life cycle and is divided into the following stages:

1. Question definition:Clear sentiment analysis goals consistent with Trump's 2025 tariff policy.

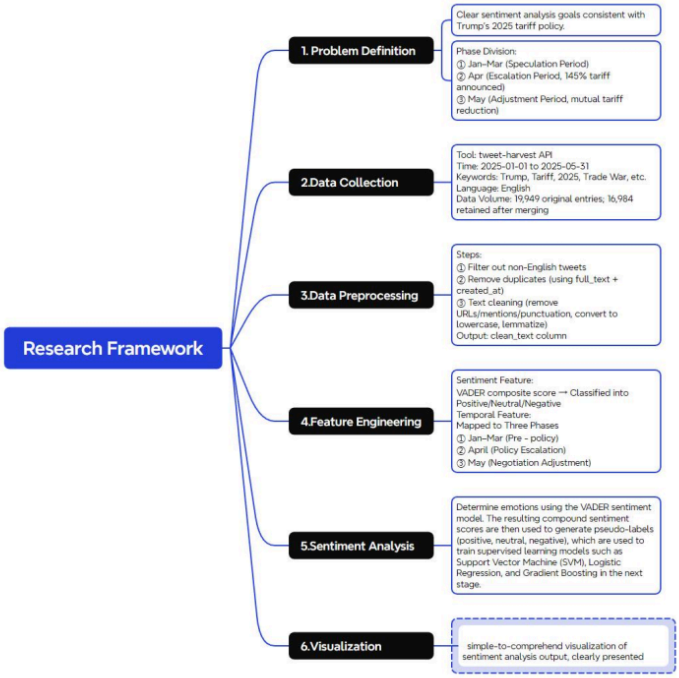
2. Data collection: Collect the significant "X" data with the appropriate keywords and in a particular time span.

3. Data preparation: Prepare the data and clean it thoroughly in order to make it more credible.

4. Feature construction: Develop features from emotions and time in a correct analysis.

5. Sentiment analysis: Determine emotions using VADER emotion model. The resulting sentiment scores are then used to generate pseudo-labels to train a supervised learning model in the next stage.

6. Visualization; simple-to-comprehend visualization of sentiment analysis output, clearly presented



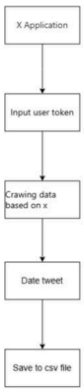
3.1 Research Framework

3.3 Problem Formulation

This research is interested in examining how individuals' perspectives are altered at three phases of policy.
January-March 2025: Retrospection on policies and initial responses
New policy adjustments and 145% tariffs in April 2025 were announced.
May 2025: Policy changes and reduced tariffs among nations.

Key objectives:
Apply VADER to determine whether tweets are positive, neutral, or negative.
Examine how individuals' attitudes shifted during three different time frames.

3.4 Data Sources & Collection

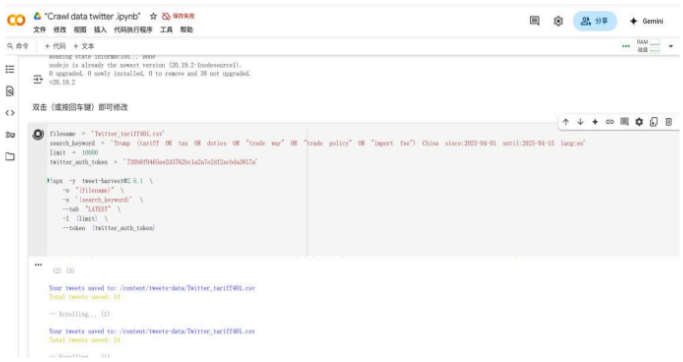


This data was collected from the Twitter (X) platform using Python-based tweet crawler tool ("Tweet harvest").

Keywords used: "Trump", "Tariff", "2025", "China", "Policies", "Trade War"

Duration: January 1, 2025 to May 31, 2025

Language filtering: Only English tweets (' lang:en ').



3.2 Data Collection

The dataset contains:

a.Tweets ("whole text")

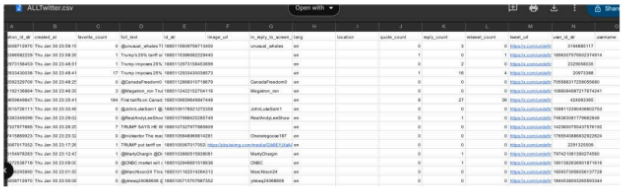
b.Timestamp ("create time")

c.Like, share, reply (for user stickiness analysis)

Combine the data mined separately every month into a data set."

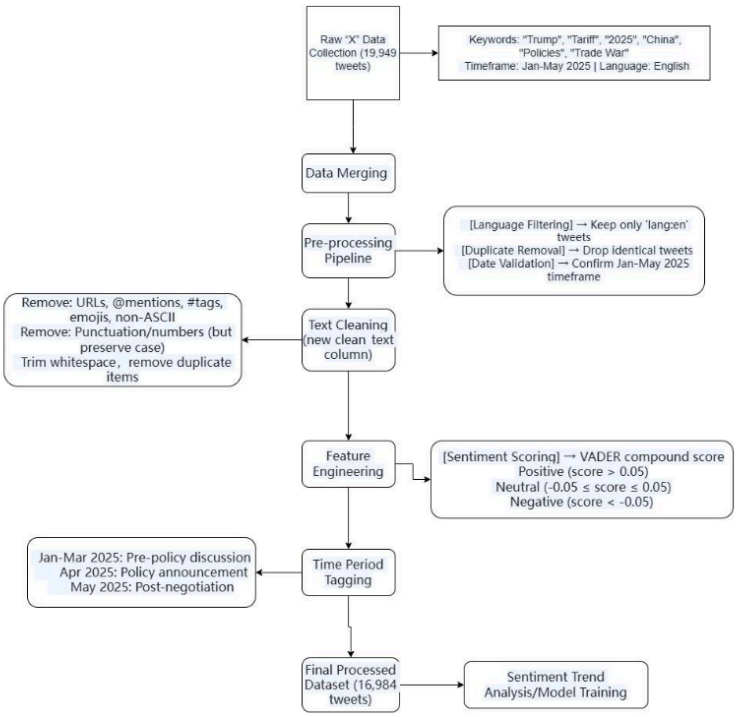
3.3 Data Merging

The total data set collected is 19,949 rows of data, including 15 columns.



3.4 Dataset preview

3.5 Data Pre-processing



Once we filtered the languages, deleted duplicates, and eliminated noise, we had 16,984 tweets. This is our final data set that we will examine in terms of sentiment scores and trends through time.

```
print(df[["created_at", "full_text", "clean_text"]].head())

                       created_at  \
0  Thu Jan 30 23:59:19 +0000 2025
1  Thu Jan 30 23:58:26 +0000 2025
2  Thu Jan 30 23:48:51 +0000 2025
3  Thu Jan 30 23:48:41 +0000 2025
4  Thu Jan 30 23:48:25 +0000 2025

                                           full_text  \
0  @unusual_whales Threat? He literally did this ...
1  Trump's 25% tariff on Canada and Mexico; a hig...
2  Trump imposes 25% tariffs on Canada and Mexico...
3  Trump imposes 25% tariffs on Canada and Mexico...
4  @CanadaFreedom0 This is why Trump is tariff fo...

                                          clean_text
0  threat he literally did this in  people wake u...
1  trumps  tariff on canada and mexico a higher o...
2  trump imposes  tariffs on canada and mexico fr...
3  trump imposes  tariffs on canada and mexico fr...
4  this is why trump is tariff focused  mexico  t...

print("Total tweets after cleaning:", len(df))

Total tweets after cleaning: 16984
```

3.5 Processed data

There are two key steps in data preprocessing. These steps ensure the data is clean, consistent, and prepared for sentiment analysis using the VADER model.

3.5.1 Preliminary analysis

- Verify that the data is correct for the target period (January to May 2025).
- There is a simple check to determine the number of tweets posted during a particular period. This is to ensure the dates listed in the policy (for instance, approximately April 10, 2025) are accurate.
- Ensure the lang field exists and is solely utilized for filtering tweets that are in English.

3.5.2 Data cleaning

To ensure that the tweets are good for the VADER model, we performed some cleaning steps. We retained all the emotional parts but got rid of typical noise in social media text.

Gradually clean the logic:

Filter English tweets

Only the English notes were preserved according to Vader's list of emotional words in the English language.

Remove duplicate data

Use the entire text and place in columns to eliminate duplicate tweets so that every comment is counted once only.

```
import pandas as pd
import re

# 读取数据文件
df = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/ALLTwitter1.csv')
df = df[df["lang"] == "en"].copy()

# 步骤 2：去除重复推文（同文本同时间被视为重复）
df = df.drop_duplicates(subset=["full_text", "created_at"])

# 步骤 3：定义清洗函数
def clean_text(text):
        text = re.sub(r"http\S+|www\.\S+", "", text)      # 删除URL
        text = re.sub(r"@\w+", "", text)                  # 删除@提及
        text = re.sub(r"#\w+", "", text)                  # 删除hashtag
        text = re.sub(r"[^\w\s]", "", text)               # 删除标点符号
        text = re.sub(r"\d+", "", text)                   # 删除数字
        text = text.lower().strip()                       # 小写化并去空格
        return text

# 步骤 4：应用清洗函数
df["clean_text"] = df["full_text"].astype(str).apply(clean_text)

# 可选：显示前几行结果检查
print(df[["created_at", "full_text", "clean_text"]].head())
```

```
                  created_at  \
0  Thu Jan 30 23:59:19 +0000 2025
1  Thu Jan 30 23:58:26 +0000 2025
2  Thu Jan 30 23:48:51 +0000 2025
3  Thu Jan 30 23:48:41 +0000 2025
4  Thu Jan 30 23:48:25 +0000 2025

                                           full_text  \
0  @unusual_whales Threat? He literally did this ...
1  Trump's 25% tariff on Canada and Mexico; a hig...
2  Trump imposes 25% tariffs on Canada and Mexico...
3  Trump imposes 25% tariffs on Canada and Mexico...
4  @CanadaFreedom0 This is why Trump is tariff fo...
```

3.6 Data cleaning

Text preprocessing

Cleaning text

1. Delete links: VADER does not comprehend links; they will be puzzling.

2. Remove mention (@user): mention feels emotional.

3. Take away tag: Tag sign is removed but the word can be retained to facilitate the understanding.

4. Eliminate emoji and non-ASCII characters: These are excluded since they are not text input.

5. Remove punctuation and numbers: Vader explains words, not symbols or numbers.

6. no conversion to lowercase: standardized text. e.g., "tariffs" is equivalent to "tariffs".

7. Trim blank areas: Remove excess space for consistency.

8. Clean storage results:

9. The cleaned result is saved in a new column called clean_text, which is used as input for sentiment analysis

3.6 Feature Engineering

Improve tweet data to analyze emotions and examine over time trends. Improve the meaning and time components. VADER is frequently employed since it is suitable for social media text (Chavan et al., 2024;) (Gandy et al., 2025), and parts of it were employed in sorting emotions.

In order to further improve the effect of sentiment classification, this study not only uses the Compound scores generated by the VADER model for trend analysis, but also uses them as pseudo-labels for model training. A balanced labeled dataset is constructed from these pseudo-labels, which will be used to train supervised learning models such as support vector machines, random forests, and logistic regression. This semi-supervised modeling strategy will be elaborated in the next chapter.

3.6.1 Emotional Feature Extraction

The cleaned tweet (clean_text) is then analyzed using the VADER sentiment analyzer in order to obtain a score representing the overall mood of a tweet.

Emotional tags are assigned according to these levels of scores:

- Positive:Compound score > 0.05
- Neutral:-0.05 ≤Compound score ≤ 0.05
- Negative: compound score <-0.05

3.6.2 Time feature generation

Each tweet is tagged with a particular date and one of three various policy times:

- January to March 2025: Discuss it and consider thoroughly before legislating.
- April 2025: Peak tariff policy announcement and reactions of other

nations.

- May 2025: Post-adjustment period after reciprocity negotiations

3.8 Results visualization

Displaying results The labeled emotional data is plotted on a chart in order to observe public opinion trends over time.

Line graph: It demonstrates how the average rating varies with different months, in relation to significant policy events.

Vertical bar chart: It displays the number of positive, negative, and neutral tweets during each phase of the policy.

Word cloud: Display typical words in various emotional groups so that people can comprehend what the public is interested in.

Python tools such as matplotlib, seaborn, and wordcloud are employed in visualization to represent emotional trends simply and simply.

These tweets are classified based on VADER's (valence-aware dictionary and emotion inference) compound score. This is suitable to classify the brief and informal messages in Twitter tweets.

3.9 Summary

This chapter shows how to set up an analysis process based on the VADER model. In the next chapter, we will further explain not only the three-stage sentiment trend, but also how VADER scores are used to generate pseudo-labels and then build a semi-supervised sentiment classification model.

CHAPTER 4

INITIAL RESULTS

4.1 Introduction

This chapter presents the preliminary results of an analysis of Trump's 2025 tariff policy on China, with a focus on outcomes and public sentiment. Firstly, the data is identified and preprocessed, including data cleaning and generation of temporal features, and the dataset is divided into three different temporal phases. Motivated by this, this chapter uses the VADER sentiment analysis tool to assess the emotional tone of the data, providing valuable insights into public reactions. In addition, the research scope is extended by constructing a supervised learning model. The proposed model uses sentiment labels generated by VADER as pseudo labels, which aims to enhance the reliability and feasibility of sentiment classification and thus gain a more detailed understanding of public sentiment surrounding tariff policies.

4.2 Exploratory Data Analysis (EDA)

EDA is a very important step in understanding your data. In this chapter, we will remove invalid, duplicate, and incorrect records during our data cleaning process to ensure data quality and consistency, and to lay a solid foundation for subsequent analysis.

For feature engineering, time features are generated, and the distribution and change rules of data in the time dimension are obtained by mining time information, which is conducive to analyzing emotional tendencies at different times.

Analyzing the length of tweets to grasp the size of words helps to grasp the richness and expressive characteristics of the content.

In visual analysis, time visualization presents the time trend of data in an intuitive chart, which facilitates the discovery of temporal patterns.

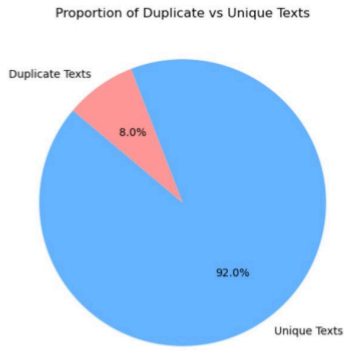
The word cloud is generated to visually display high-frequency words, reflect the core topics and concerns of the data, and provide clues for the depth mining of data information.

In preparation for sentiment analysis, the VADER model is initialized to provide a powerful tool for subsequent sentiment analysis.

Looking at the compound score distribution gives us a preliminary picture of the overall sentiment orientation, paving the way for sentiment analysis and modeling.

4.2.1 Data cleaning generates temporal features

After data cleaning and deduplication, the final data is 16984 data sets, in which the proportion of repeated values is 8% and the proportion of unique values is 92%. The temporal features are generated according to the months, which are the three time periods of January-March, April, and May.



4.1 Proportion of Duplicate vs Unique Texts

```
[4]: #生成时间特征#
     # 强制转换 created_at 为 datetime 类型，指定格式更安全
     df['created_at'] = pd.to_datetime(df['created_at'], format='%a %b %d %H:%M:%S %z %Y', errors='coerce')

     # 过滤掉转换失败的行（即created_at为空的）
     df = df[df['created_at'].notna()].copy()

     # 添加月份字段
     df['month'] = df['created_at'].dt.to_period('M').astype(str)

     # 添加政策阶段字段
     def get_policy_period(date):
         if date < pd.to_datetime("2025-04-01", utc=True):
             return "Jan-Mar"
         elif date < pd.to_datetime("2025-05-01", utc=True):
             return "April"
         else:
             return "May"

     df['policy_period'] = df['created_at'].apply(get_policy_period)

     # 检查
     print(df[['created_at', 'month', 'policy_period']].head())


C:\Users\holy\AppData\Local\Temp\ipykernel_22580\73087119.py:9: UserWarning: Converting to PeriodArray/Index
on.
  df['month'] = df['created_at'].dt.to_period('M').astype(str)
                 created_at    month policy_period
0 2025-01-30 23:59:19+00:00  2025-01       Jan-Mar
1 2025-01-30 23:58:26+00:00  2025-01       Jan-Mar
2 2025-01-30 23:48:51+00:00  2025-01       Jan-Mar
3 2025-01-30 23:48:41+00:00  2025-01       Jan-Mar
4 2025-01-30 23:48:25+00:00  2025-01       Jan-Mar
```

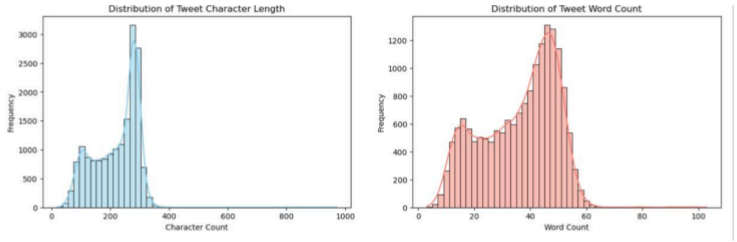
4.2Generation time feature

4.2.2 Tweet length word count analysis

Through visual analysis, the core features of tweets are revealed:

(a) Character length: In most tweets, it is 100-300 characters, with a peak of about 200 characters, showing a long-tail distribution (the proportion of tweets with the number of characters > 400 decreases sharply), reflecting the propagation characteristics of short texts.

(b) Number of words: The number of words is concentrated in the range of 20 to 60 words, with a peak of about 40 words, which is consistent with the trend of character length, and verifies the ecological characteristics of "refined expression" in tweets.



4.3 Length word count analysis

It reflects the characteristics of short text and long-tail distribution, and provides a basis for the subsequent design of sentiment analysis model：

4.2.3 The number of tweets was counted by month

The figure shows the monthly tweet count trend from January to May 2025, with month on the horizontal axis and number of tweets on the vertical axis.

(a) Foundation stage (Jan.-Feb.) : Low discussion

From January to early February 2025, the Trump administration did not officially launch the tariff policy against China, and the public discussion on the policy gradually decreased.

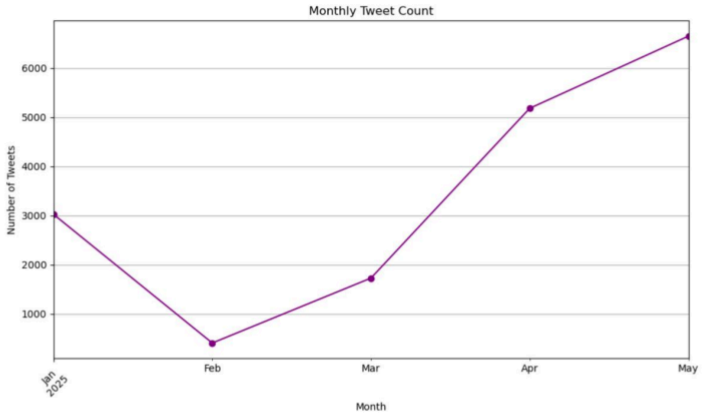
(b) Rising period (February-April) : Discussions gradually increase

During this period, Trump signed an executive order announcing a 10% tariff on imports from China, and the discussion rate rose and the slope was steep. After hitting bottom in February, the discussion rate gradually increased from March to April (March ≈1800→ April ≈5200), marking the beginning of the "discussion rate rising period" :

(c) Apri-May: Discussion degree peak breakthrough

April-may 2025; President Trump signed a "reciprocal tariff" executive order, raising tariffs on China to 125%

The growth trend continued from April to May, when it reached its annual peak (over 6,500) and entered a "discussion explosion" :



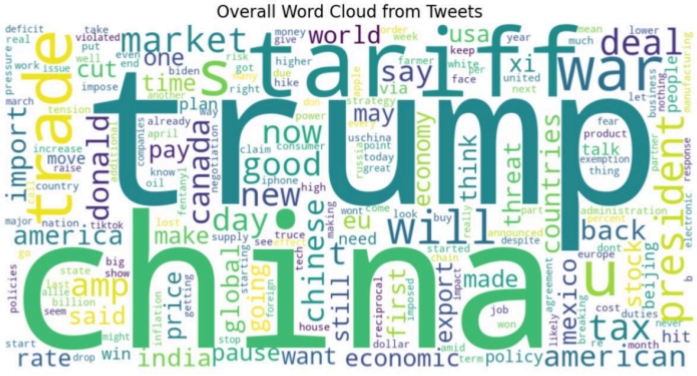
4.4 Monthly Tweet Count

4.2.4 Word cloud map

The word graphs of the generated data are consistent with the topics, and high-frequency words are not only aligned with the core topics (trade, tariffs, and China-US interaction), but also imply sentiment tendencies and sub-topic branches. Based on this, sentiment classification can be performed in the later stage.

Among them：

- Conflict and game: Words such as "war", "threat" and "risk" reflect discussions of trade conflicts and policy threats in tweets, suggesting tensions caused by tariff policies;
- economy and market: The words "market", "economy", "price" and "product" reflect a focus on the economic impact of tariff policies (market fluctuations, cost changes);
- Action and response: "impose", "cut", "plan", "talk", etc., denoting discussion of policy implementation actions and response strategies (such as tariff imposition and trade negotiations)。



4.5 Overall Word Cloud from Tweets

4.3 Sentiment analysis

4.3.1 compound sentiment score distribution

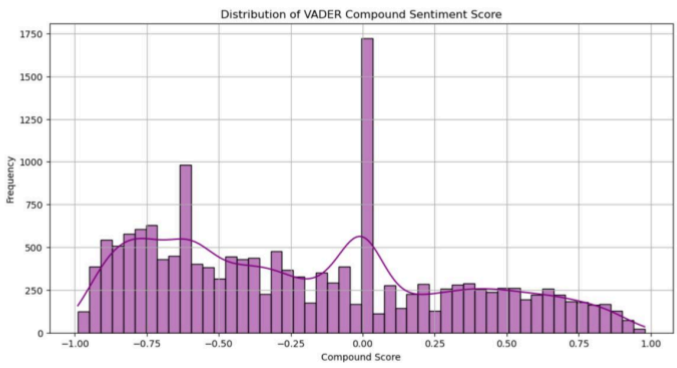
Based on the VADER sentiment analysis tool, the sentiment labels are generated and the composite sentiment score distribution is calculated.

(a) Score distribution features

- There is a clear neutral tendency: the highest frequency of the composite score is around 0 (neutral range), indicating that neutral sentiment accounts for the largest proportion of tweets
- Polarization exists: scores around -1(very negative) and 1(very positive) are less common, but there is a clear peak between -0.75 and -0.5(negative range), suggesting that negative sentiment discussions should not be ignored

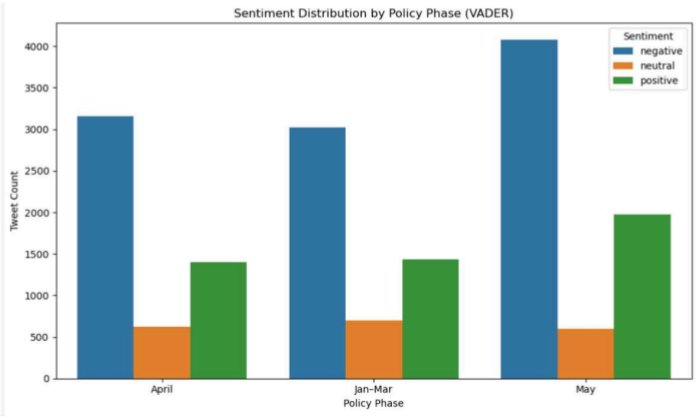
(b) Connect

- Sentiment tone verification: neutral sentiment is dominant, which is consistent with the characteristics of "policy discussion type text".
- Sentiment analysis direction: Negative sentiment peaks are the focus of mining objects
- Model Applicability: The VADER score distribution shows a "high in the middle, low at both ends" shape, indicating that the tool performs well on neutral text recognition on this dataset.

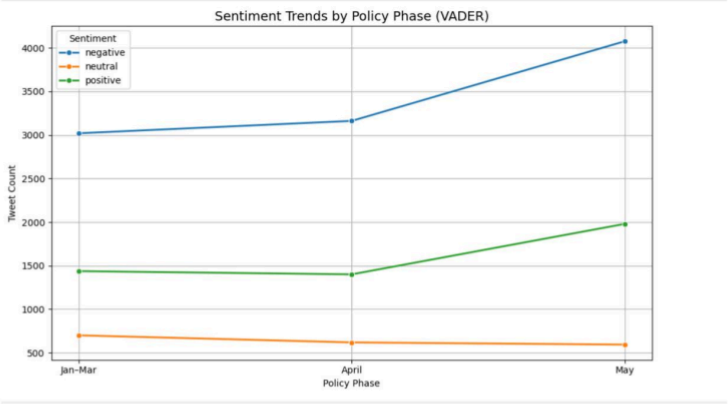


4.6 Distribution of VADER Compound Sentiment Score

4.3.2 Count emotions in stages



4.7 Sentiment Distribution



4.8 Sentiment Trends

(a) January-march (Policy fermentation period: tax increase starts in February-March)

Policy background: 10% tariff was imposed in February, and the tax exemption policy was cancelled. In March, the tariff was raised to 20 percent, the first round of retaliatory measures (Chinese tariffs on American agricultural products and American tariffs on steel and aluminum).

Emotional characteristics:

Bar: Highest proportion of negative sentiment (≈3100) and lower proportion of neutral (≈700) and positive (≈1400).

Line chart: negative sentiment stable, positive, neutral no significant fluctuations.

Analysis: The policy has been initially implemented, the public opinion is dominated by "worrying about the impact", supporters and opponents have not yet formed fierce confrontation, and neutral content still has room for survival.

(b) April (extreme confrontation period: the tax rate soared to 125% in April)

Policy background: In April, "reciprocal tariffs" were superimposed (a comprehensive tax rate of 145%), and China and the US retaliated with additional tariffs (China's rare earth control, US chips and other goods were exempted).

Emotional characteristics:

Bar chart: Negative sentiment increased slightly (≈3200 bars), positive sentiment increased substantially (≈1450 bars), and neutral sentiment was flat (≈700 bars).

Line chart: Negative sentiment starts to accelerate, positive sentiment bottoms out, neutral continues to decline.

Analysis: Extreme policies activate "public opinion confrontation" :

Negative sentiment among opponents is exacerbated by "a high tax rate of 125% and the risk of supply chain disruption";

Supporters actively voice, positive emotions rise;

Rational discussion (neutral) is squeezed by "different positions", and the proportion continues to decline.

(c) Temporary tariff cuts in May + negotiations

Policy background: Tariffs were temporarily lowered to 30%(US)/10%(China) in May, entering the negotiation window.

Emotional characteristics:

Bar chart: negative sentiment peaks (≈4100), positive sentiment peaks at the same time (≈2000), and neutral sentiment bottoms (≈600).

Line chart: Negative sentiment increases rapidly, positive sentiment continues to rise, and neutral drops to a minimum.

Analysis: The "temporary alleviation" after extreme confrontation triggers the "emotional surplus + expectation game" :

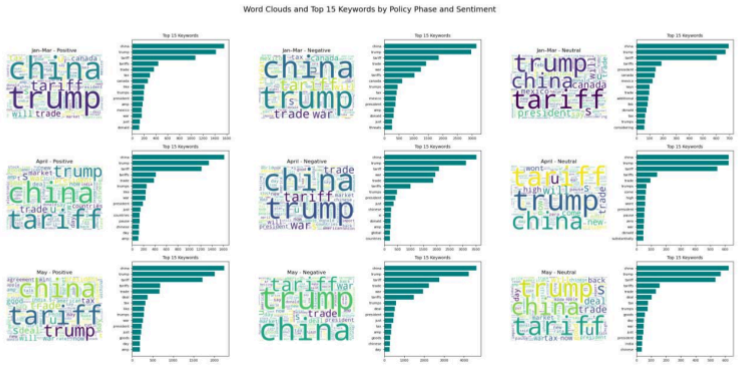
(a) ●Opponents dissatisfied with the "repeated policy", negative feelings hit a record high;

(b) ●Supporters see "negotiation" as a signal of policy victory, and positive sentiment continues to rise;

(c) ●Prolonged confrontation makes public opinion highly emotional, and neutral content is almost marginalized.

In short, negative sentiment follows the tariff "bungee jump" : the higher the tariff, the greater the sarcasm; Even after the temporary tariff reduction in May, concerns about the economy and supply chain simmered, with negative waves hitting new highs. The more confrontational the policy, the more intense the "supporters' quarrel" : the beneficiaries of trade protection and ethnic sentiment groups are "lit on fire" by extreme policies, and the one-way attention is turned into a melee of insults on both sides. Rational analysis is completely "hidden" : from the beginning of the year to May and June, the voice of objective facts and analysis diminishes. After April and May, there was hardly any mention of it. It was all opposition。

4.3.3 Emotional high-frequency word analysis at different stages



4.9 High-frequency words of emotions at different stages

Jan-Mar (Policy fermentation period: the first round of tax increases/countermeasures in February-March)

(a)Positive

• High-frequency words: china, tariff, trump, will, trade

• Logic: Proponents are optimistic about the effect of the policy, such as "will" implies the expectation of the deterrence of tariffs (the belief that higher tariffs will force China to compromise); trade focuses on trade games and reflects the "trade protectionism" stance.

(b) Negative

• High frequency words: china, tariff, trump, trade war

• Logic: Opponents fear the escalation of conflict, and "trade war" directly points to the fear of "trade war outbreak" (the first high-intensity confrontation between the two sides due to the tax increase in February and the countermeasure in March).

(c) Neutral

• High frequency words: trump, china, mexico, tariff, president

• Logic: objectively stating policy subjects and actions, such as "president says" recording Trump's tariff statement, without obvious emotional bias.

April (extreme confrontation period: the tax rate soared to 125% in April)

(a)Positive

High-frequency words :china, tariff, trump, deal, market

Logic: Proponents see "extreme tax increases" as a bargaining chip, and "deal" suggests they expect tough policies to force China to sign a favorable deal; The market may point to "domestic market protection" (such as manufacturing interests).

(b)Negative

High-frequency words :china, tariff, trump, trade war, market

Logic: Opponents focus on economic shocks, while "market" highlights concerns about "market turbulence and supply chain disruptions" (April's 125% tax rate hits business costs directly); The trade war escalated into reality and negative sentiment intensified.

(c) Neutral

High-frequency words :tariff, trump, china, will, come

Logic: Keep an objective record of the policy process. For example, "tariff will come" describes the April 5 arrival date equivalent to a 125% tariff.

May (temporary grace period: tariff reduction in May+negotiation in May)

(a) Positive

High-frequency words: China, tariffs, Trump, deals, agreements and benefits.

Logic: Supporters recognize the breakthrough in the negotiations, and "agreement/agreement" reflects a positive interpretation of "temporary tariff reduction and negotiation window period" (regarded as policy victory); Good directly expresses the affirmation of the result.

(b)Negative

High-frequency words: tariff, Trump, China, trade war, trade.

Logic: Opponents question the value of the agreement, and "agreement" means dissatisfaction with "temporary tax reduction and negotiation compromise" (such as thinking that there are too many concessions or worrying about policy duplication); The trade war is still going on, indicating that people are worried about the recurrence of the conflict.

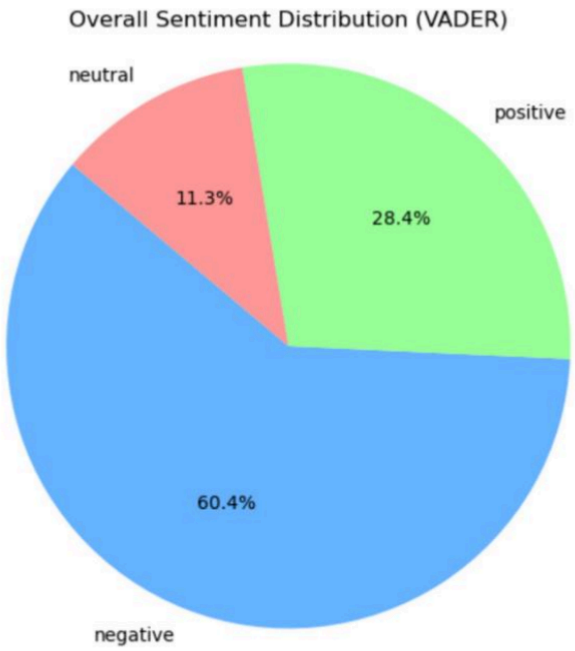
(c)Neutral

High-frequency words: Trump, China, tariff, market, transaction.

Logic: objectively state the result of the event, such as "transaction" only refers to the "China-US tariff agreement" itself, without emotion, focusing on the factual record of policy adjustment.

sentiment dimension	**Jan-Mar (fermentation)**	**April (confrontation)**	**May (relaxation)**
positive	Look forward to the policy "will"	Looking forward to the "deal"	Recognize the "agreement"
negative	Worried about the "trade war"	Worried about "market"	Query the "deal"
Neutral	Record "president"	Record "Policy Landing" (will/come)	Record the "deal"

4.3.4 Overview of public opinion and emotion under tariff policy



4.10 Overall Sentiment Distribution

Overall tone

Negative emotions account for 60.4%, positive emotions account for 28.4%, and neutral emotions account for 11.3%, less than 40%, which confirms the negative feedback of public opinion caused by tariff policy. "

Development stage

Jan-mar (policy fermentation): the negative is dominant (≈60%), which stems from the concern about the "impact of tax increase"; The positive reason is that "expecting policy deterrence" (such as "tariffs will change the trade pattern") has a certain weight; A neutral record policy statement (such as "Trump Statement").

April (extreme confrontation): negative acceleration (≈65%), because the ultra-high tax rate of 125% detonated the "fear of economic shock" (such as supply chain rupture); Positive because "tough gambling agreements" (such as "tax increase forcing China to make concessions") have increased; Neutral compression is a "policy landing record" (such as "the effective time of tariffs").

May (temporary relief): the negative peak (≈70%) stems from "repeated distrust of policies" (such as "whether temporary tax cuts can be sustained"); On the positive side, "optimistic negotiation results" (such as "reaching an agreement") increased slightly, but still weak; There is only "agreement fact statement" in neutrality, and rational discussion is completely marginalized.

Emotional logic driven by policy

The rhythm of Trump's tariff on China "confrontation and escalation → temporary relaxation" directly shaped the feeling of public opinion:

The more radical the policy is (for example, the tax rate of 125% in April), the stronger the negative sentiment and the stronger the voice of confrontation (for example, the supporters of trade protection advocate high tariffs);

The policy has turned to relaxation (such as tax reduction in May), the negative residue is still stubborn (afraid of "policy duplication"), and the neutral space has been completely squeezed.

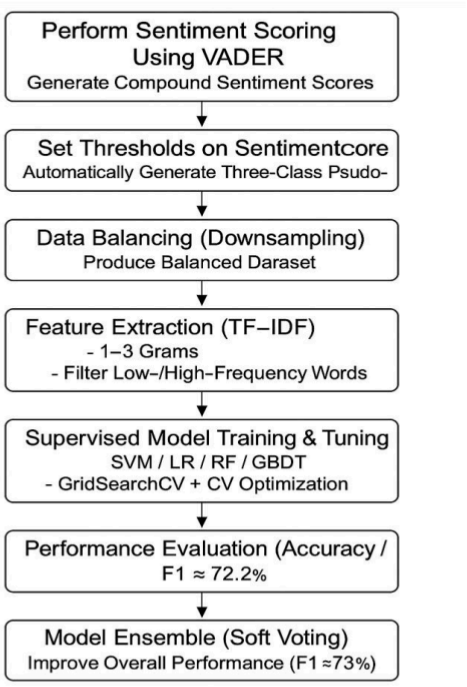
Finally, the law of "the stronger the policy antagonism, the more serious the polarization of public opinion" is formed.

In a word, the "tough-compromise" cycle of tariff policy makes public opinion fall into the predicament of "negative dominance, bipolar confrontation and rational

loss", and the negative proportion of 60.4% is the emotional price of trade policy conflict.

4.4 Extended analysis of emotion classification model based on VADER pseudo-label

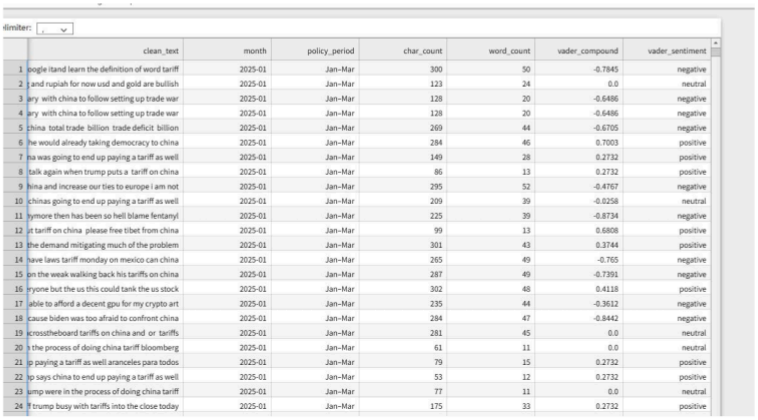
In addition to the unsupervised sentiment analysis based on VADER, a supervised learning model is further tried to be constructed, and the traditional classifier is trained by using the sentiment polarity result output by VADER as a "pseudo-label" to evaluate its classification reliability and modeling feasibility.。



4.11 Supervised sentiment classification model construction process

4.4.1 Pseudo-label Generation

In order to quickly generate sentiment label to train the supervision model without manual tagging, this study calculated the comprehensive emotional score of tariff policy-related tweets in 2025 with the help of VADER tool specially designed for social media texts, and automatically divided them into positive, neutral and negative emotions as pseudo-tags.



	clean_text	month	policy_period	char_count	word_count	vader_compound	vader_sentiment
1	oogle itand learn the definition of word tariff	2025-01	Jan-Mar	300	50	-0.7845	negative
2	and rupiah for now usd and gold are bullish	2025-01	Jan-Mar	123	24	0.0	neutral
3	ary with china to follow setting up trade war	2025-01	Jan-Mar	128	20	-0.6486	negative
4	ary with china to follow setting up trade war	2025-01	Jan-Mar	128	20	-0.6486	negative
5	china total trade billion trade deficit billion	2025-01	Jan-Mar	269	44	-0.6705	negative
6	he would already taking democracy to china	2025-01	Jan-Mar	284	46	0.7003	positive
7	na was going to end up paying a tariff as well	2025-01	Jan-Mar	149	28	0.2732	positive
8	talk again when trump puts a tariff on china	2025-01	Jan-Mar	86	13	0.2732	positive
9	hina and increase our ties to europe i am not	2025-01	Jan-Mar	295	52	-0.4767	negative
10	chinas going to end up paying a tariff as well	2025-01	Jan-Mar	209	39	-0.0258	neutral
11	ymore then has been so hell blame fentanyl	2025-01	Jan-Mar	225	39	-0.8734	negative
12	t tariff on china please free tibet from china	2025-01	Jan-Mar	99	13	0.6808	positive
13	the demand mitigating much of the problem	2025-01	Jan-Mar	301	43	0.3744	positive
14	have laws tariff monday on mexico can china	2025-01	Jan-Mar	265	49	-0.765	negative
15	on the weak walking back his tariffs on china	2025-01	Jan-Mar	287	49	-0.7391	negative
16	ryone but the us this could tank the us stock	2025-01	Jan-Mar	302	48	0.4118	positive
17	able to afford a decent gpu for my crypto art	2025-01	Jan-Mar	235	44	-0.3612	negative
18	cause biden was too afraid to confront china	2025-01	Jan-Mar	284	47	-0.8442	negative
19	crosstheboard tariffs on china and or tariffs	2025-01	Jan-Mar	281	45	0.0	neutral
20	the process of doing china tariff bloomberg	2025-01	Jan-Mar	61	11	0.0	neutral
21	p paying a tariff as well aranceles para todos	2025-01	Jan-Mar	79	15	0.2732	positive
22	p says china to end up paying a tariff as well	2025-01	Jan-Mar	53	12	0.2732	positive
23	ump were in the process of doing china tariff	2025-01	Jan-Mar	77	11	0.0	neutral
24	f trump busy with tariffs into the close today	2025-01	Jan-Mar	175	33	0.2732	positive

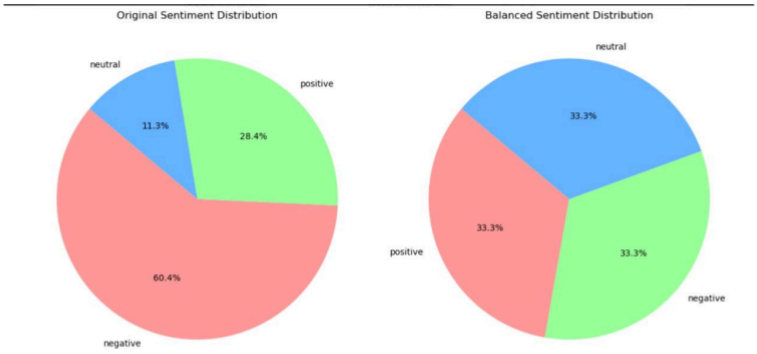
4.12 Generation of pseudo-labels

4.4.2 Data balance processing

Because VADER's output is unbalanced (for example, the proportion of negative samples is too high), a balanced data set (balanced _ VADER _ sentinel. Construct CSV by down sampling) to ensure the fairness of model learning.

```
import pandas as pd
from sklearn.utils import resample
import matplotlib.pyplot as plt

# 1. 加载原始数据（包含 clean_text 列的完整数据）
df = pd.read_csv('vader_phase_sentiment.csv')  # 替换为实际文件路径

# === 原始数据集情绪分布 ===
original_counts = df['vader_sentiment'].value_counts()

# 设定目标样本数量（最小类 Neutral 的数量）
target_size = original_counts.min()

# 2. 修改下采样逻辑：确保保留所有列
balanced_df = pd.DataFrame()
for sentiment in ['positive', 'negative', 'neutral']:
    subset = df[df['vader_sentiment'] == sentiment]
    resampled = resample(subset, replace=False, n_samples=target_size, random_state=42)
    balanced_df = pd.concat([balanced_df, resampled])

# 现在检查 balanced_df 的列，应该包含 clean_text 了
print("balanced_df 的列：", balanced_df.columns.tolist())

# === 对比原始 vs 平衡后分布 ===
fig, axes = plt.subplots(1, 2, figsize=(12, 6))

# 原始分布饼图
axes[0].pie(
    original_counts,
    labels=original_counts.index,
    autopct='%1.1f%%',
    startangle=140,
    colors=['#ff9999', '#99ff99', '#66b3ff']
)
axes[0].set_title("Original Sentiment Distribution")

# 平衡分布饼图
balanced_counts = balanced_df['vader_sentiment'].value_counts()
axes[1].pie(
    balanced_counts,
    labels=balanced_counts.index,
    autopct='%1.1f%%',
    startangle=140,
    colors=['#ff9999', '#99ff99', '#66b3ff']
)
axes[1].set_title("Balanced Sentiment Distribution")

plt.tight_layout()
plt.show()
```

4.13 Balanced dataset



4.14 Balanced dataset pie chart

```
原始分布：
 vader_sentiment
negative    10255
positive     4816
neutral      1913
Name: count, dtype: int64

平衡后分布：
 vader_sentiment
negative    1913
positive    1913
neutral     1913
Name: count, dtype: int64
```

4.15The quantity after balancing the data

4.4.3 Data Enhancement Pretreatment

Before the model training, we strengthened the preprocessing of tweet text, and considered "noise filtering" and "emotional information retention":

Noise filtering and emotional punctuation:

First, delete URL, user mention (@), subject tag (#) and numbers, and keep punctuation marks related to emotion (! ?) and standardize the format (continuous punctuation marks are separated by spaces, such as huge! ! ! Think of it as huge! !), which not only reduces the irrelevant interference, but also retains the emotional strength.

Text standardization process:

(d) ●Lowercase: unify vocabulary case (such as tariff → tariff);

(e) ●word segmentation: divide the text into lexical units;

(f) ●stop words filtering: removing high-frequency meaningless words (such as and);

(g) ●word form restoration: restore the vocabulary to the basic form (such as tariff → tariff) to ensure that different word forms are mapped to the same features.

Role:

After preprocessing, tweets are transformed into standardized vocabulary sequences, which can eliminate irrelevant noise and preserve emotional semantics, laying a foundation for subsequent TF-IDF feature extraction and model training.

```python
# 2. 增强数据预处理
print("\n增强文本预处理...")
lemmatizer = WordNetLemmatizer()
stop_words = set(stopwords.words('english'))

def enhanced_text_preprocessing(text):
    """增强的文本预处理函数"""
    if not isinstance(text, str):
        return ""

    # 基本清理
    text = re.sub(r'http\S+|@\w+|#\w+|\d+', '', text)  # 移除URL、提及、标签和数字
    text = re.sub(r'[^\w\s]', ' ', text)  # 移除非字母数字字符（保留空格）

    # 保留情感相关的标点（如! ?）
    text = re.sub(r'([!?])', r' \1 ', text)  # 给标点加空格

    # 词形还原和过滤
    tokens = nltk.word_tokenize(text.lower())
    tokens = [lemmatizer.lemmatize(token) for token in tokens
              if token not in stop_words and len(token) > 1]

    return " ".join(tokens)
```

4.16 Data enhancement preprocessing

4.4.4 Feature Extraction

Core goal: to transform tweets into numerical features that focus on policy semantics and strengthen emotional differentiation, which is suitable for emotional analysis of short texts.

N-gram coverage:

Extract 1-3 yuan phrases (such as tariffs, trade wars and new tariff policies) to obtain complete semantic units related to policies;

Feature space optimization:

Retain 8000 terms with the largest amount of information (filtering rare words and super-commonly used words) to balance the model efficiency and semantic coverage.

Noise control:

Filter words that appear less than 3 times (min_df=3) and words that exceed 70% of documents (max_df=0.7), and focus on the core vocabulary of policy discussion.

Weight balance:

Logarithmically scale the word frequency (subblinear_tf =True) to avoid the excessive dominance of high-frequency words (such as tariff) and highlight the role of emotional phrases (such as terrible tariff).

By preserving emotional punctuation (for example!) Combined with TF-IDF, it strengthens the capture of strategic emotional intensity and provides accurate input for the emotional classification model.

```
# 创建特征提取器（TF-IDF + 词袋）
tfidf = TfidfVectorizer(
    max_features=8000,   # 增加特征数量
    ngram_range=(1, 3),  # 包含一元、二元和三元语法
    min_df=3,            # 忽略低频词
    max_df=0.7,          # 忽略高频词
    sublinear_tf=True,   # 使用对数TF
    stop_words='english'
)

# 特征降维（可选）
svd = TruncatedSVD(n_components=1000, random_state=42)
```

4.17 TF-IDF

4.4.5 Supervised model training and hyperparameter optimization

Then, several classification models such as logistic regression, support vector machine (SVM), random forest and Gradient Boosting are trained and evaluated. This model uses GridSearchCV and 3 fold cross validation for hyperparameter tuning.

```python
# 5. 模型选择与优化
print("\n模型训练与优化...")

# 定义更优的模型和参数网格
models = {
    'Logistic Regression': {
        'model': LogisticRegression(max_iter=2000, random_state=42, class_weight='balanced'),
        'params': {
            'clf__C': [0.01, 0.1, 1, 10],
            'clf__solver': ['saga', 'liblinear'],
            'clf__penalty': ['l1', 'l2']
        }
    },
    'SVM': {
        'model': SVC(probability=True, random_state=42, class_weight='balanced'),
        'params': {
            'clf__C': [0.1, 1, 10],
            'clf__kernel': ['linear', 'rbf'],
            'clf__gamma': ['scale', 'auto']
        }
    },
    'Random Forest': {
        'model': RandomForestClassifier(n_estimators=200, random_state=42, class_weight='balanced_subsample'),
        'params': {
            'clf__max_depth': [None, 30, 50],
            'clf__min_samples_split': [2, 5, 10],
            'clf__min_samples_leaf': [1, 2, 4]
        }
    },
    'Gradient Boosting': {
        'model': GradientBoostingClassifier(n_estimators=200, random_state=42),
        'params': {
            'clf__learning_rate': [0.01, 0.1],
            'clf__max_depth': [3, 5],
            'clf__subsample': [0.8, 1.0]
        }
    }
}

# 存储评估结果
results = {}
best_models = {}
```

4.18 Model selection and optimizati

```python
# 使用分层K折交叉验证
cv = StratifiedKFold(n_splits=3, shuffle=True, random_state=42)

for name, model_info in models.items():
    print(f"\n=== 优化 {name} ===")

    # 创建管道
    pipeline = Pipeline([
        ('tfidf', tfidf),
        # ('svd', svd),  # 如果需要降维可以启用
        ('clf', model_info['model'])
    ])

    # 网格搜索
    grid_search = GridSearchCV(
        pipeline,
        model_info['params'],
        cv=cv,
        scoring='accuracy',
        n_jobs=-1,  # 使用所有核心
        verbose=1
    )

    grid_search.fit(X_train, y_train)

    # 获取最佳模型
    best_model = grid_search.best_estimator_
    best_params = grid_search.best_params_

    # 评估测试集
    y_pred = best_model.predict(X_test)
    acc = accuracy_score(y_test, y_pred)
    f1 = f1_score(y_test, y_pred, average='weighted')

    print(f"{name} 最佳参数: {best_params}")
    print(f"{name} 测试集准确率: {acc:.4f}")
    print(f"{name} 测试集F1值: {f1:.4f}")
    print(classification_report(y_test, y_pred))

    # 保存结果
    results[name] = {
        'model': best_model,
        'accuracy': acc,
        'f1': f1,
        'params': best_params,
        'report': classification_report(y_test, y_pred, output_dict=True)
    }
    best_models[name] = best_model
```

4.19 K-fold cross validation

4.4.6 Model comparison and evaluation

After completing the model training and parameter optimization, we use the test set with emotion tags generated by VADER who did not participate in the training, and evaluate the performance of logistic regression, SVM, random forest and Gradient Boosting with accuracy and F1-score (macro average or weighted average of three categories) as indicators.

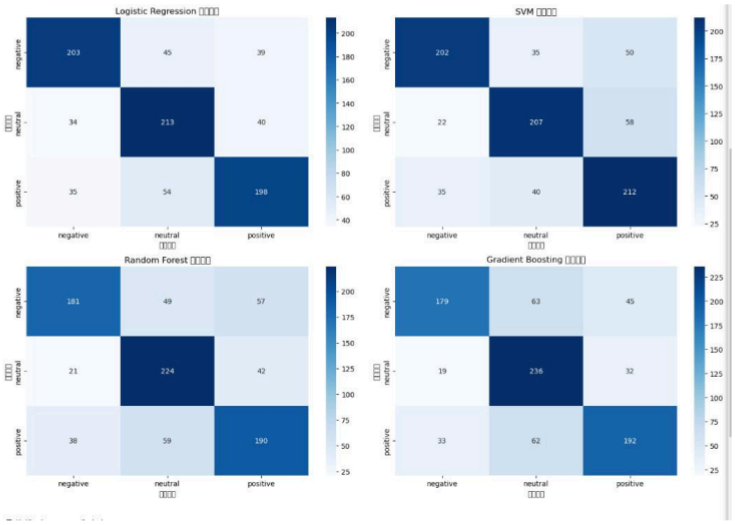
Model name	**Accuracy**	**F1 Score**	**Optimal superparameter configuration**
SVM	0.721254	0.722012	clf_C: 10, clf_gamma: 'scale', clf_kernel: 'rbf'
Logistic Regression	0.713124	0.713203	clf_C: 10, clf_penalty: 'l1', clf_solver: 'liblinear'
Gradient Boosting	0.704994	0.703387	clf_learning_rate: 0.1, clf_max_depth: 5, clf_subsample: 1.0
Random Forest	0.691057	0.690126	clf_max_depth: None, clf_min_samples_leaf: 1, clf_min_samples_split: 10

	Model	Accuracy	F1 Score
1	SVM	0.721254	0.722012
0	Logistic Regression	0.713124	0.713203
3	Gradient Boosting	0.704994	0.703387
2	Random Forest	0.691057	0.690126

4.20 Comparison of Models

In a single model, SVM and logistic regression performed best, and SVM was slightly better than logistic regression. The accuracy and F1 value of decision tree-

based models (random forest and gradient boosting model) are relatively low, which may be related to the limited data set size or the noise of false labels.



4.21 Confusion matrix

4.4.7 integration model

Soft Voting Ensemble Classifier (integrating the best logistic regression, SVM and gradient boosting model) achieves the best overall performance by combining the strongest single model, with an accuracy of about 73.05% and a F1 value of about 73.00%, which exceeds the performance of any single model. This shows that the integrated model can take advantage of the complementary advantages of different classifiers, thus improving the prediction performance.

```python
# 7. 最佳模型选择与保存
best_model_name = model_comparison.iloc[0]['Model']
best_model = results[best_model_name]['model']
best_accuracy = results[best_model_name]['accuracy']

print(f"\n最佳模型: {best_model_name} (准确率: {best_accuracy:.4f})")

# 保存最佳模型
joblib.dump(best_model, 'best_sentiment_model.pkl')
print("最佳模型已保存为 'best_sentiment_model.pkl'")

# 8. 集成模型（可选）
print("\n尝试集成模型...")
from sklearn.ensemble import VotingClassifier

# 选择前2-3个最佳模型进行集成
top_models = model_comparison.head(3)['Model'].tolist()
estimators = [(name, best_models[name]) for name in top_models]

voting_clf = VotingClassifier(
    estimators=estimators,
    voting='soft',  # 使用概率投票
    n_jobs=-1
)

voting_clf.fit(X_train, y_train)
y_pred_voting = voting_clf.predict(X_test)
acc_voting = accuracy_score(y_test, y_pred_voting)

print(f"集成模型准确率: {acc_voting:.4f}")
print(classification_report(y_test, y_pred_voting))

# 如果集成模型表现更好，则保存它
if acc_voting > best_accuracy:
    joblib.dump(voting_clf, 'best_ensemble_model.pkl')
    print("集成模型表现更佳，已保存为 'best_ensemble_model.pkl'")

print("\n优化完成!")
```

4.22 Model integration

```
最佳模型: SVM (准确率: 0.7213)
最佳模型已保存为 'best_sentiment_model.pkl'

尝试集成模型...
集成模型准确率: 0.7305
              precision    recall  f1-score   support

    negative       0.76      0.71      0.74       287
     neutral       0.71      0.76      0.74       287
    positive       0.72      0.72      0.72       287

    accuracy                           0.73       861
   macro avg       0.73      0.73      0.73       861
weighted avg       0.73      0.73      0.73       861

集成模型表现更佳，已保存为 'best_ensemble_model.pkl'

优化完成!
```

4.23 Model integration complete

4.4.8 Advantages and limitations of the method

Advantages: VADER's emotional understanding of social texts is effectively utilized, and training data is obtained at a lower cost; Through balance processing and model integration, the accuracy of VADER is compensated.

Limitations: The accuracy rate is limited by the quality of pseudo-labels (the essence of the model is to reproduce Vader's judgment), and the accuracy rate of 73% is not ideal; Pure automation process is difficult to deal with complex semantics (such as irony and fuzzy expression), and it needs iterative optimization through a small amount of manual annotation (active learning).

In a word, the extended experiment shows the idea of constructing pseudo-tag training set from unsupervised results and refining the model through supervised learning, which embodies the feasibility of semi-supervised learning in social media sentiment analysis. Cheng Kewei's subsequent construction of a model with higher performance and more interpretability provides a method reference, and also verifies the practicability of VADER in the initial emotional screening stage.

4.5 Summary

This chapter takes the social media texts under the background of Trump's tariff policy towards China in 2025 as the research object, and completely presents the analysis process from data exploration to emotional modeling:

Data exploration stage: mining the time trend, text characteristics and word frequency distribution of tweets through EDA, and showing the evolution of public attention through word borrowing cloud and phased tweets.

Emotional analysis stage: using VADER tool to score unsupervised emotions, it is found that emotional tendencies are "caused by negative emotions, rising polarization and marginalized rational voices".

Model construction stage: build a pseudo-label balanced data set by VADER score, and extract features by Term Frequency–Inverse Document Frequency (TF-IDF); After training and optimization, it is found that SVM performs best, and the accuracy and F1 value of the integrated model are improved to 73%, which verifies the effectiveness of the "pseudo-label+integrated supervision model".

This study lays a foundation for the follow-up discussion of public opinion mechanism and policy response simulation, shows the quantitative path of emotional evolution of social media, and highlights the application potential of semi-supervised learning in policy-sensitive public opinion analysis.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1 Introduction

In order to further summarize the key results of this study and explore its theoretical and practical significance, this chapter summarizes the research results and puts forward some suggestions for future research. In the previous chapter, based on VADER's emotional analysis and pseudo-label supervised learning, we constructed a set of social media emotional modeling process, revealing how Trump's tariff policy towards China in 2025 triggered significant fluctuations in public opinion emotions at different stages. This chapter first summarizes the phased core findings and structural changes, and then puts forward the potential limitations of the current method and the feasible improvement direction, in order to provide reference for future related research.

5.2 Conclusion

This study focuses on the emotional analysis of social media under the background of Trump administration's tariff policy towards China in 2025 (January-May), and constructs a complete research process from data exploration to model construction. The core findings are as follows:

a) The overall emotional mode is negative, and the positive and negative polarization is obvious:

According to VADER's analysis, about 60.4% of tweets are negative, much higher than positive (28.4%) and neutral (11.3%), which reflects the public's concern and aversion to high-intensity trade policies.

Public opinion mood fluctuates obviously with the policy stage:

- Jan-March (policy fermentation period): public opinion is dominated

by "wait and see+worry", and negative emotions appear at first, but there is still room for discussion between positive and neutral;

- April (period of extreme confrontation): The sharp increase in tariffs to 125% triggered an emotional outbreak and negative emotions continued to rise. At the same time, there are obvious support voices, positive emotions rise, and polarization of public opinion begins to appear;
- May (temporary easing period): Although the policy turned to easing (tariff reduction+negotiation) in the short term, due to the public's distrust of "policy duplication", the negative emotions did not decrease but increased, reaching the peak of the whole cycle, and the neutral rational content was marginalized.

(h)

b) The stage analysis reveals that:

The higher the resilience of the policy, the more extreme the mood; Even in the easing period, the negative "residual emotions" have not dissipated, indicating that the credibility and consistency of policies have a far-reaching impact on public psychology.

High-frequency words reveal the focus of public attention and emotional drivers: positive words focus on negotiation-oriented words such as "transaction" and "agreement", negative words focus on words related to economic shocks such as "trade war" and "market", and neutral words are mostly policy statements, which further reflects the close combination of emotion and policy narrative.

A semi-supervised learning path based on pseudo-tags is proposed: pseudo-tags are automatically generated by VADER's emotional score, a balanced data set is constructed, text features are extracted by TF-IDF, and SVM, logistic regression and random forest models are trained respectively. In the end, the accuracy of the integrated model of soft voting reached 73.05%, which was better than any single model, and verified the effectiveness of the strategy of "pseudo-label+integrated learning" under the condition of lack of manual labeling.

c) Summary of research contribution:

This study has made the following contributions from two dimensions: methodology and demonstration:

- It systematically reveals the evolution path of "tariff policy → change of public opinion structure → emotional polarization";
- Propose the strategy of "Vader pseudo-label+supervised learning" without manual labeling to expand the practicability of emotion modeling;
- The quantitative coupling mechanism between emotional expression and policy influence in social media platform is verified, which provides theoretical and methodological support for future policy communication and public response prediction.

5.3 Future Work

Although the emotional analysis and set model of this study have some findings, there are still some limitations, and the following optimization directions can be considered in future research:

a) The influence of pseudo-label noise on model performance.

Because the emotional tags generated automatically by VADER may misjudge complex semantics (such as irony, fuzziness and ambiguity), the tags are inaccurate and the upper limit of the model is limited.

Suggestion: Active learning or self-training mechanism can be introduced in combination with a small amount of manual labeling to improve the accuracy of pseudo-labels and the generalization ability of models.

b) Lack of deep semantic modeling ability

This study adopts shallow features such as TF-IDF, which is efficient and interpretable, but it is difficult to capture contextual and syntactic information and identify deep meaning.

Suggestion: Integrate pre-training language models (such as BERT and RoBERTa) and introduce Transformer structure to improve the recognition ability of complex emotions such as implied semantics, irony and pun.

c) Single platform and language source

This study only analyzes the English text of Twitter, but does not cover other platforms (such as Reddit and Facebook) and other language expressions, which limits the scope of generalization.

Suggestion: Expand multi-platform and multi-language public opinion analysis, explore the similarities and differences of responses in different cultural backgrounds, and enhance the breadth and applicability of research.

5.4 Summary

To sum up, this paper systematically reveals the emotional evolution mechanism of social media triggered by different stages of Trump's tariff policy towards China in 2025, and proposes a semi-supervised modeling method based on pseudo-tags. It is found that public opinion fluctuates significantly with the evolution of policy intensity, showing a dynamic pattern of "negative dominance-bipolar opposition-lack of rationality". This study not only verifies the applicability of VADER in political public opinion scenes, but also improves the performance of emotion classification through model integration, demonstrating the feasibility of low-cost public opinion modeling.

Although there are limitations in label quality, model depth and data source, the proposed process still provides a theoretical framework and technical path for future policy public opinion analysis, social media emotion tracking and automatic risk early warning, which has certain practical value and research expansion potential. Improve the breadth and applicability of research.

References

Chavan, R., Latthe, S., Dhorepati, M., Suryawanshi, A., Sharma, N., & Salge, A. (2024). Sentiment analysis using VADER & word cloud techniques. AIP Conference Proceedings, 3217(1), Article 020012. AIP Publishing. https://doi.org/10.1063/5.0234543.

ElMassry, A. M., Alshamsi, A., Abdulhameed, A. F., Zaki, N., & Belkacem, A. N. (2024). Machine learning approaches for sentiment analysis on balanced and unbalanced datasets. Proceedings of the 2024 IEEE 14th International Conference on Control System, Computing and Engineering (ICCSCE), 18–23.

Gandy, L. M., Ivanitskaya, L. V., Bacon, L. L., & Bizri-Baryak, R. (2025). Public health discussions on social media: Evaluating automated sentiment analysis methods. JMIR Formative Research, 9, e57395. https://doi.org/10.2196/57395

Guo, S., Jiao, Y., & Xu, Z. (2021). Trump's effect on the Chinese stock market. Journal of Asian Economics, 72, 101267. https://doi.org/10.1016/j.asieco.2020.101267

Perumal Chockalingam, S., & Thambusamy, V. (2024). Enhancing sentiment analysis of user response for COVID-19 vaccinations tweets using SentiWordNet-adjusted VADER sentiment analysis (SAVSA): A hybrid approach. In K. Iyakutti, P. Balasubramaniam, & K. R. Subramanian (Eds.), Lecture Notes in Networks and Systems: Vol. 1046. Proceedings of the International Conference on Recent Advances in Computational Techniques (IC-RACT 2024) (pp. 437–451). Springer. https://doi.org/10.1007/978-3-031-64813-7_43

Pham, D. P. T., Huynh, N. Q. A., & Duong, D. (2022). The impact of US presidents on market returns: Evidence from Trump's tweets. Research in International Business and Finance, 62, Article 101681.

Selmi, R., Errami, Y., & Wohar, M. E. (2020). What Trump's China tariffs have cost U.S. companies? Journal of Economic Integration, 35(2), 282–295. https://doi.org/10.11130/jei.2020.35.2.282

Son, M. (2022). The global propagation of the US–China trade war. Empirical Economics, 63(6), 3121–3157. https://doi.org/10.1007/s00181-022-02231-7

Wengerek, S. T., Uhde, A., & Hippert, B. (2025). Share price reactions to tariff imposition announcements during the first Trump administration. Finance Research Letters, 80, Article 107381. https://doi.org/10.1016/j.frl.2025.107381

Zangmo, D., Dar, A. I., Kumar, R., & Mishra, V. N. (2024). Sentiment analysis on U.S. election. AIP Conference Proceedings, 3005(1), 020024. https://doi.org/10.1063/5.0210583

Zheng, J., Zhou, S., Li, X., Padula, A. D., & Martin, W. (2023). Effects of eliminating the US-China trade dispute tariffs. World Trade Review, 22(2), 212–231. https://doi.org/10.1017/S1474745622000271

Gjerstad, P., Meyn, P. F., Molnár, P., & Næss, T. D. (2021). Do President Trump's tweets affect financial markets? Decision Support Systems, 147, 113577. https://doi.org/10.1016/j.dss.2021.113577

Loynes, C., Ouenniche, J., & De Smedt, J. (2022). Detection and estimation of disaster locations using Twitter and identification of NGOs using crowdsourcing. Annals of Operations Research, 308, 339–371.

Wengerek, S. T., Uhde, A., & Hippert, B. (2025). Share price reactions to tariff imposition announcements during the first Trump administration. Finance Research Letters, 80, 107381. https://doi.org/10.1016/j.frl.2025.107381

Dwianto, R. A., Nurmandi, A., & Salahudin, S. (2021). The sentiments analysis of Donald Trump and Jokowi's Twitters on COVID-19 policy dissemination. Webology, 18(1), 389–405. https://doi.org/10.14704/WEB/V18I1/WEB18096

Pham, D. P. T., Huynh, N. Q. A., & Duong, D. (2022). The impact of US presidents on market returns: Evidence from Trump's tweets. Research in International Business and Finance, 62, 101681. https://doi.org/10.1016/j.ribaf.2022.101681

Abdollahi, H., Fjesme, S. L., & Sirnes, E. (2024). Measuring market volatility connectedness to media sentiment. The North American Journal of Economics and Finance, 71, 102091. https://doi.org/10.1016/j.najef.2024.102091

Nishimura, Y., & Sun, B. (2025). Impacts of Donald Trump's tweets on volatilities in the European stock markets. Finance Research Letters, 72, 106491. https://doi.org/10.1016/j.frl.2024.106491

Zhang, Q., Frömmel, M., & Baidoo, E. (2024). Donald Trump's tweets, political value judgment, and the Renminbi exchange rate. International Review of Financial Analysis, 93, 103159. https://doi.org/10.1016/j.irfa.2024.103159

Faridzi, Z. A., Pramesti, D., & Fa'Rifah, R. Y. (2023). A comparison of oversampling and undersampling methods in sentiment analysis regarding Indonesia fuel price increase using support vector machine. In Proceedings of ICADEIS 2023—International Conference on Advancement in Data Science, E-Learning and Information Systems: Data, Intelligent Systems, and the Applications for Human Life (pp. not provided). https://doi.org/10.1109/ICADEIS58666.2023.10270851

Zangmo, D., Dar, A. I., Kumar, R., & Mishra, V. N. (2024). Sentimental analysis on U.S. election. AIP Conference Proceedings, 3005(1), 020024. https://doi.org/10.1063/5.0210583

Thesis_LI HONGLIN.pdf

ORIGINALITY REPORT

10%	9%	2%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universiti Teknologi Malaysia Student Paper	6%
2	www.coursehero.com Internet Source	1%
3	link.springer.com Internet Source	<1%
4	ntnuopen.ntnu.no Internet Source	<1%
5	www.readkong.com Internet Source	<1%
6	repo.lib.tokushima-u.ac.jp Internet Source	<1%
7	www.americaspg.com Internet Source	<1%
8	Submitted to Universiti Teknikal Malaysia Melaka Student Paper	<1%
9	brentnixon.github.io Internet Source	<1%

10	www2.mdpi.com Internet Source	<1%
11	Submitted to Queen Mary and Westfield College Student Paper	<1%
12	pmc.ncbi.nlm.nih.gov Internet Source	<1%
13	Submitted to Bocconi University Student Paper	<1%
14	learnsql.com Internet Source	<1%
15	www.diplomarbeiten24.de Internet Source	<1%
16	Sascha Tobias Wengerek, André Uhde, Benjamin Hippert. "Share price reactions to tariff imposition announcements during the first Trump administration", Finance Research Letters, 2025 Publication	<1%
17	digitalscholarship.unlv.edu Internet Source	<1%
18	Ahli, Hajar Hussain. "Sales Opportunities Lead Qualification in B2B Market.", Rochester Institute of Technology Publication	<1%

19	dataknowsall.com Internet Source	<1%
20	dfrws.org Internet Source	<1%
21	discovery.researcher.life Internet Source	<1%
22	dr.library.brocku.ca Internet Source	<1%
23	portfolios.cs.earlham.edu Internet Source	<1%
24	repository.tudelft.nl Internet Source	<1%
25	text-id.123dok.com Internet Source	<1%
26	www.igi-global.com Internet Source	<1%
27	www.scss.tcd.ie Internet Source	<1%
28	Ernest N. Biktimirov, Tatyana Sokolyk, Anteneh Ayanso. "Unpacking the Relation Between Media Sentiment and House Prices: A Topic Modeling Approach", Journal of Housing Economics, 2024 Publication	<1%

29	"Artificial Intelligence Applications and Innovations. AIAI 2025 IFIP WG 12.5 International Workshops", Springer Science and Business Media LLC, 2025 Publication	<1%
30	"Intelligent Systems Design and Applications", Springer Science and Business Media LLC, 2024 Publication	<1%
31	Mohammad Ashraful Ferdous Chowdhury, Mohammad Abdullah, Mousa Albashrawi. "Analyzing public sentiment toward economic stimulus using natural language processing", Transforming Government: People, Process and Policy, 2024 Publication	<1%
32	Peder Gjerstad, Peter Filip Meyn, Peter Molnár, Thomas Dowling Næss. "Do President Trump's tweets affect financial markets?", Decision Support Systems, 2021 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		