Title

Personality-Driven Sentiment: Linking Myers-Briggs Type Indicator (MBTI) Types to Emotional Expression on Social Media

Author

Luo, Yuchen

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Personality-Driven Sentiment: Linking Myers-Briggs Type Indicator (MBTI) Types to Emotional Expression on Social Media

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Applied Statistics and Data Science

by

Yuchen Luo

2024

ABSTRACT OF THE THESIS

Personality-Driven Sentiment:
Linking Myers-Briggs Type Indicator (MBTI) Types
to Emotional Expression on Social Media

by

Yuchen Luo
Master of Applied Statistics and Data Science
University of California, Los Angeles, 2024
Professor Yingnian Wu, Chair

The Myers-Briggs Type Indicator (MBTI) is a widely popular personality classification tool. While being used for personal growth, it has also gradually aroused strong public interest in the field of social media. However, there has been a lack of clear and quantitative evidence on whether different MBTI personality types have significant differences in emotional expression on social media. This article aims to explore this issue through rigorous data processing and sentiment analysis. The data set comes from Kaggle's MBTI data set. We clean the data and retain text features, and use the VADER analyzer to perform sentiment analysis on the language style of social media users. At the same time, we use a variety of visualization methods to compare the emotional distribution and word usage characteristics of different MBTI types. We then identify underlying patterns through intuitive visualizations.

The results show that people with extroversion (E) and feeling (F) types are more likely to express positive emotions on social media, while people with introversion (I) and thinking (T) types are relatively neutral. These findings help to better understand the dynamic relationship between personality characteristics and social media behavior, and also provide inspiration for personalized recommendations, marketing strategies, psychology research and other fields.

The thesis of Yuchen Luo is approved.

Frederic R. Paik Schoenberg

Nicolas Christou

Yingnian Wu, Committee Chair

University of California, Los Angeles

2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Thanks to everyone I met at UCLA. Your kindness and support made my journey of study warm and fulfilling. Special thanks to my parents—your selfless dedication and unwavering support have given me the opportunity to stand on your shoulders and see a broader world.

CHAPTER 1

Introduction

Personality type represents a framework for classifying different individual based on their characteristic ways of thinking and behavior. It differentiates people to distinct categories. Also, it provides a systematic method for interpreting the broad difference in human cognition and behavior. Its application can be applied to several domains, including psychology, organizational behavior, education, etc. The core purpose of personality type is to better understand different ways of people's interaction, communication, and pursue their personal or professional targets.

1.1 Introduction of Myers-Briggs Type Indicator (MBTI)

The Myers-Briggs Type Indicator (MBTI) is one of the most well-known tools that describes personality. It was created by Katharine Cook Briggs and Isabel Briggs Myers. The MBTI divides people into one of sixteen types. Each of MBTI type is defined by a combination of four key pairs of traits: Extraversion (E) or Introversion (I), Sensing (S) or Intuition (N), Thinking (T) or Feeling (F), and Judging (J) or Perceiving (P).

Each pair shows a double side of different personality. Extraversion (E) or Introversion (I) focuses on whether someone gets energy from communicating with others

like social event or from staying alone. In general, people with Extraversion personality prefer social activities and get energy from them, while people with Introversion personality usually stay at home to restore their energy after a busy day. Sensing (S) or Intuition (I) is about whether someone pays more attention to facts and details or to abstract ideas and concepts. Thinking (T) or Feeling (F) focuses on whether they base decisions mostly on logic or on emotions. Judging (J) or Perceiving (P) is about whether someone prefers a more planned approach or a more open and flexible time schedule. By combining these preferences, the MBTI describes sixteen types that share certain patterns in how they interact with others and solve problems. And many of professional portals, such as 16Personalities [16P] and MyPersonality [Per] divide these sixteen personalities into four roles.

1.2 The Sixteen MBTI Personality Types

1. ISTJ: People with this type are often practical and responsible. These folks often have a logical perspective on life and are restrained but assertive. They meticulously plan their activities and execute them with deliberate intent.

2. ISFJ: ISFJs usually show kindness and thoughtfulness. They often put other people's needs before their own. They are careful with details and try to help others feel safe and at ease. They like traditions and routines. Their caring nature makes them good listeners and trustworthy friends or coworkers.

3. INFJ: INFJs care about people and want to make a positive difference. They often sense what others feel and think about what the future might bring. They usually approach life with a great deal of imagination and thinking. They are

guided in everything by their inner vision, personal principles, and a subdued, moral form of humanism.

4. INTJ: INTJs are often logical and good at seeing patterns or long-term possibilities. They like clear ideas and careful thinking. They often set high standards for themselves and others. These strategic thinkers like fine-tuning the little things in life, bringing creativity and reason to all they undertake. They frequently have a complicated and secretive inner life.

5. ISTP: ISTPs often enjoy working with their hands practically. And they like solving problems they can see right in front of them. They are good at dealing with new or sudden situations. They tend to have an individualistic mindset, pursuing goals without needing much external connection. They engage in life with inquisitiveness and personal skill, varying their approach as needed.

6. ISFP: ISFPs tend to be gentle, kind and sensitive. They like to investigate about beauty and artistic conception. They often embrace life, new experiences, and people with a grounded warmth and an open mind. They discover fascinating possibilities thanks to their capacity for present-moment awareness.

7. INFP: INFPs think a lot about their values and look for deeper meanings in life. They feel strong care and concern for people and causes they believe in. Quiet, open-minded, and creative, with these uncommon personality types, they approach everything with compassion and originality.

8. INTP: INTPs rely on logic and enjoy new ideas and theories. These adaptable thinkers take pleasure in approaching many facets of life in an unusual way.

They frequently look for unusual routes, combining their own inventiveness with a desire to try new things.

9. ESTP: ESTPs are active and like to face problems directly. They pay attention to what is happening right now and try to fix things right away. They are usually vivacious and focused on taking action, skillfully handling any situation that comes their way. Whether interacting with people or engaging in more alone activities, they enjoy discovering life's possibilities.

10. ESFP: ESFPs are usually lively and enjoy being around people. They like hands-on activities and focus on the present moment. These people like having exciting experiences, living life to the fullest, and enjoying the unknown. They frequently encourage people to participate in shared activities and may be quite gregarious.

11. ENFP: ENFPs love to explore new ideas and solutions. They care about making genuine connections with other people. They often take interesting actions. At the same time, they show sincerity and kindness to others. They are full of energy and have many new ideas; they have many different path to develop.

12. ENTP: ENTPs enjoy testing ideas through debate and questioning. They are brave and imaginative, analyzing and reorganizing concepts in unique ways. They pursue their goals aggressively, even in the face of obstacles.

13. ESTJ: ESTJs feel comfortable with organizing tasks, setting rules, and making sure things work properly. They are quite resilient and firmly adhere to their own sound judgment. Among other things, they frequently act as a stabilizing factor, providing firm guidance in the face of difficulty.

14. ESFJ: ESFJs are warm, caring, and enjoy helping others feel included. They pay attention to people's feelings and try to keep good relationships. They are responsible and good at handling everyday tasks. By giving support and encouragement, they help groups work together better.

15. ENFJ: ENFJs want to encourage people to do their best. They often understand what others need and try to bring everyone together to reach common goals. They like to be organized and take on leadership roles. Their friendly guidance helps others feel motivated and valued.

16. ENTJ: ENTJs are confident, plan ahead, and use logic to achieve their aims. They want to improve systems and find better ways of doing things. They set high standards, both for themselves and the people around them. They like to take charge and may not hesitate to make tough decisions if it helps move things forward.

1.3 Four MBTI Roles

1. Analysts are intuitive and thinking (NT) personality types: INTJ, INTP, ENTJ, and ENTP. Characterized by logic, creativity, and creativity, they are passionate about anything that inspires thinking. They not only enjoy deep conversations, but also like to solve complex problems that require unconventional solutions. They have distinct personalities, strong curiosity, and a desire to constantly improve themselves through exploration and learning. Because they value rationality and strategy, Analysts tend to express their opinions and make decisions based on facts and data (such as quantitative evidence).

They can remain calm and objective in any situation, making fair judgments at critical moments.

2. Diplomats are intuitive and feeling (NF) personality types: INFJ, INFP, ENFJ, and ENFP. Diplomats are full of idealism and compassion, and always have a vision to make the world a better place. They truly care about others and fight against injustice. At the same time, they also focus on exploring the deeper meaning of life, so they are often attracted to participate in public welfare activities such as charity and volunteering. Compared to other personality types, Diplomats are less attracted to career success and material wealth, and they desire deeper emotional connections with others. In their minds, success means being able to actively help the lives of others. Because of this, and their natural creativity, many Diplomats are keen on writing, art, and other forms of artistic expression.

3. Sentinels are sensing and judging (SJ) personality types, which include ISTJ, ISFJ, ESTJ, and ESFJ. Sentinels attach great importance to stable and predictable order. They believe in rules and think that rules can make society more organized and make everyone's life better. As hardworking and reliable Sentinels, they always complete tasks conscientiously and strictly demand themselves. They are amazingly active and rarely touch things that have no practical significance. Because they are cautious and love to plan, they are sometimes reluctant to accept changes. Unless it is really necessary, they prefer to stay in their familiar comfort zone.

4. Explorers are sensing and perceiving (SP) personality types: ISTP, ISFP, ESFP, and ESTP. Explorers are spontaneous and adventurous. They enjoy

life and the present moment to the fullest and seize every opportunity. They are bold and would rather make mistakes than have regrets. They often step out of their comfort zone and face the unknown in order to try new things. At the same time, they are particularly sensitive to their senses and their surroundings, which allows them to discover many unnoticed beauties in life and quickly learn practical skills. They are also highly adaptable to changes and quick to act, which often plays a key role in emergencies.

1.4 The Importance of Studying MBTI

Exploring MBTI can increase our understanding of interpersonal communication, problem-solving approaches, and decision-making processes. It can provide insights into the collaborative efficacy of certain individuals. Additionally, it highlights the types of activities they may work proficiently. For instance, in a team, understanding each member's MBTI type and different MBTI's characteristic helps facilitate mutual comprehension of individual styles. This may mitigate misconceptions and foster improved collaboration, eventually increase efficiency.

On an individual level, MBTI can enhance self-awareness. Understanding different MBTI types may shed light on why individuals gravitate toward specific hobbies. It may also explain why certain tasks seem less challenging for some people. This might facilitate their personal development and find their own advantages. They may get a deeper understanding of their talents and limitations. Therefore, they will learn better to collaborate with other individuals in a more equitable manner.

1.5 MBTI's Influence in Hiring and Career Development

Nowadays, some employers use MBTI as a tool when they hire new employees. They also use MBTI to consider about who might fit for a certain job. Some companies will directly ask applicants for their MBTI type on their application portal. For example, a job that needs careful planning and logical thinking might fit people who prefer Intuition and Thinking. A job that requires friendly customer contact might suit someone who is more social and sympathy.

This does not mean that MBTI alone decides who gets a job. Instead, it can give employers extra information. It can help them build teams has more comfort atmosphere for every team member. This might make the workplace run more smoothly and efficiently. What's more, it can help employees feel more comfortable and valued. By taking concern with MBTI types, companies can also plan training programs that help people grow in different roles that match their strengths.

1.6 MBTI in Sentiment Analysis

During the Covid-19 epidemic, people have transferred many social networking and daily communication needs from real life to the Internet. After the end of Covid-19, the popularity of social media is continuing to increase, and people use social media more frequently. Kepios's analysis [Dat] indicates that, as of early October 2024, there were 5.22 billion social media users globally, or 63.8 percent of the entire world population. This signifies an augmentation of 256 million users in the preceding year, indicating an annual growth rate of 5.2%. The popularity of social media and its huge user base provide rich data resources for studying MBTI. Different

MBTIs may have different speaking styles on social media and may have different emotional expression tones. By analyzing these differences, we can better understand the behavioral characteristics of different MBTI personalities on social media, and can also provide new research directions for other related fields, such as organizational behavior.

In order to further understand and study these issues, this thesis will mainly study one main questions: Do users with different MBTI personalities show significant emotional tendencies on social media? For example, are certain types more likely to post positive or negative content?

This study will first introduce the social media dataset used for analysis, including its background, characteristics, and pre-processing methods; next, the thesis will introduce the specific steps and technical implementation of sentiment analysis; After visualizing the results, we will get some results and conclusion with a direct and visible way.

CHAPTER 2

Dataset Overview

This chapter outlines our data sources, provides data samples, and systematically explains how to clean the data and extract key information from the text.

2.1 Dataset Description

This study utilizes the Myers-Briggs Personality Type Dataset from Kaggle[J17]. This dataset is sourced from a website focused on personality discussions called "Personality Cafe" (https://www.personalitycafe.com/). This website resembles the Reddit forum and functions as a forum platform. This forum website offers a social networking platform for anyone interested in personality, facilitating discussions on all aspects of the subject inside the topic. This dataset comprises several language samples from social media users, offering academics a resource for examining the correlation between social language expression and the 16 MBTI personality types.

The dataset contains 8,675 records, with each record represented as a row containing two columns: a user's MBTI personality type and their previous post. The first column is the user's MBTI personality type, which consists of four dimensions:

Introversion (I)/Extroversion (E)

Intuition (N)/Feeling (S)

Thinking (T)/Feeling (F)

Judgment (J)/Perception (P)

The subsequent column comprises a user's 50 most recent post records on social media, delineated by $'|||'$. These writings illustrate the user's linguistic style, perspectives, and emotions across many contexts. The post encompasses a broad spectrum, perhaps addressing daily life, material dissemination, and responses or viewpoints about certain subjects. The post record may include emoticons, typographical errors, online links, and certain internet jargon. This dataset employs authentic content shared by people on social media to illustrate the authenticity and variety of social media language.

The following table is a sample with ten rows of the dataset. Each row includes the user's MBTI personality type and their social media post recordings (the length is truncated in the example presentation); nevertheless, as shown in the table, the post records may have emoticons, typographical errors, and site links. Consequently, it is essential to preprocess the data prior to initiating sentiment analysis, conduct fundamental cleaning and filtering, and preserve the necessary information required in the following analysis.

Type	Posts
ENTP	'I'm finding the lack of me in these posts very alarming.|||Sex can be boring ...
INTP	'Good one ___ https://www.youtube.com/watch?v=ft...
INTJ	'Dear INTP, I enjoyed our conversation the other day. ...
ENTJ	'You're fired.|||That's another silly misconception. That a ...
INTJ	'18/37 @.@|||Science is not perfect. No scientist claims ...
INFJ	'No, I can't draw on my own nails (haha). Those were done ...
INTJ	'I tend to build up a collection of things on my desktop that ...
INFJ	'I'm not sure, that's a good question. The distinction between ...
INTP	'https://www.youtube.com/watch?v=w8-egj0y8Qs|||I'm ...
INFJ	'One time my parents were fighting over my dad's affair a ...
INTJ	'Fair enough, if that's how you want to look at it. Lik ...
INTP	'Basically this... https://youtu.be/1pH5c1JkhLU|||I C ...
INTP	'Your comment screams INTJ, bro. Especially the use ...
INFJ	'some of these both excite and calm me: BUTTS bo ...
INFP	'I think we do agree. I personally don't consider myse ...
INFJ	'I fully believe in the power of being a protector, to g ...
INFP	'That's normal, it happens also to me. If I am in high ...

Table 2.1: Sample of the Raw Data from MBTI Dataset

2.2 Data Pre-Processing

In the data preprocessing stage, the original dataframe was systematically cleaned to improve the effectiveness of subsequent model training. The main steps are as

follows: First, regular expressions were used to match and remove all URL links in the text (such as strings starting with http:// or https://). By cleaning all URL links, we can remove noise data that is irrelevant to the text semantics and may cause model misjudgment.

To further reduce the noise in the data frame, I only retained English letters, numbers, and spaces. At the same time, I retained common punctuation marks that can help the model recognize emotions and tone (such as periods, commas, exclamation points, question marks, semicolons, and colons). In addition, I retained the existence of the vertical bar symbol —, which is used to split posts in the next step. Through this process, emoticons, special graphic characters, and other meaningless symbols are effectively removed.

At the same time, the format of the words is also strictly processed. First, all the text is converted to lowercase format. This step can reduce the feature redundancy problem caused by inconsistent capitalization of words. For example, "Love" and "love" are ultimately considered the same lexical feature; at the same time, we use WordNetLemmatizer to restore each word to its basic form. For example, "running" is restored to "run". This process reduces the impact of word singularity and pluralization, and word tense, allowing the model to understand the text with a more focused word feature representation. Secondly, letters that are repeated too many times in a row are simplified, such as "loooove" is standardized as "love". This method can reduce ambiguous words caused by users expressing their emotions strongly or input errors, and can also enhance the standardization of the text.

If the dataset contains clear MBTI type terms (such as "infp", "entj", etc.), they will be removed. This prevents the model from relying on these labels for self-

identification during training, rather than truly learning text features and human expressions. At the same time, we use stop_word in the nltk library to delete stop words in the text. Stop words appear frequently in the text, but they contribute little to the main meaning of the text. In natural language processing, stop words are usually not used to express key information. Therefore, removing them can reduce noise and make the data clearer.

Finally, we separate different posts in the original data using the built-in separator symbol "|||", and finally merge them back into a noise-free text sequence to ensure that the data still retains certain logical paragraph information after cleaning.

2.3 Exploratory Data Analysis (EDA)

During the data pre-processing phase, we refined and standardized the raw data to guarantee quality and consistency. Pre-processing enables the acquisition of a somewhat organized dataset with minimal noise, hence furnishing more solid foundational data to assist our forthcoming Exploratory Data Analysis (EDA). This part will utilize data visualization to examine the data, intuitively illustrating the distribution features of various MBTI types and prospective user behavior traits.

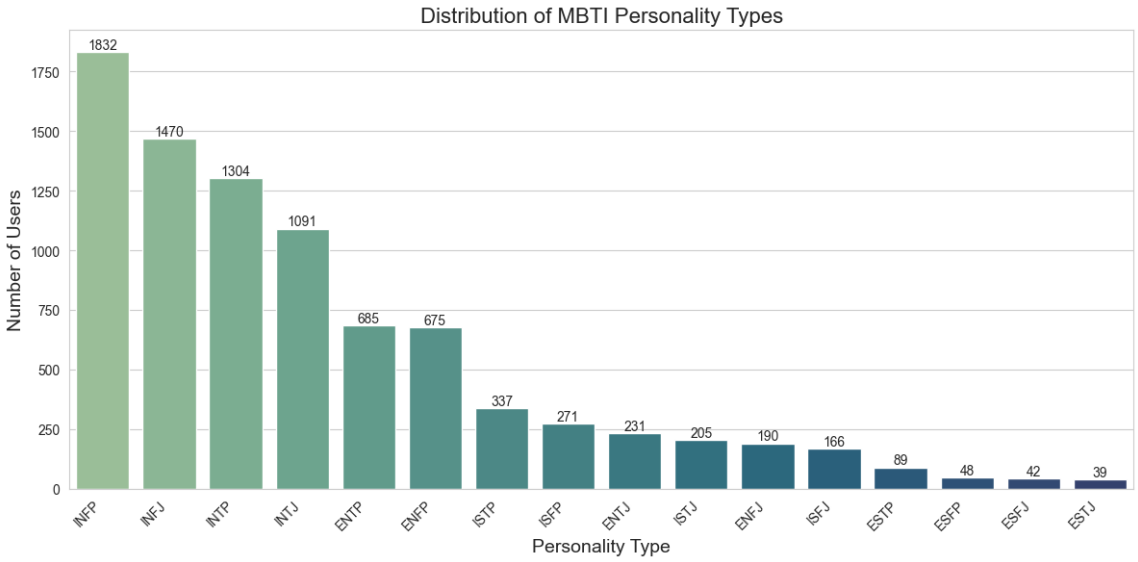

Figure 2.1: Bar Chart: Distribution of MBTI Personality Types

The bar chart (Figure 2.1: Distribution of MBTI Personality Types) clearly shows the distribution of MBTI personality types of users in the dataset. From the sorted bar chart, we can see that INFP is the most common personality type, with 1,832 people, followed by INFJ (1,470 people) and INTP (1,304 people). In contrast, the number of users of ESTP, ESFP, ESFJ and ESTJ is very small, all less than

100 people. This distribution shows that in this dataset, the number of users with intuitive personality is much larger than the number of users with sensing personality. The first five users with personality distribution are all intuitive users, while the last five users with personality distribution are all sensing users. The dataset comes from an online forum that focuses on discussing personality types. This forum is more likely to attract users who are interested in abstract concepts, which just fits the tendency of the "intuitive (N type)" personality. Users with intuitive personality are more likely to participate in conversations involving abstract theories or in-depth discussions, which may lead to a significant increase in the number of N type users in the dataset. In contrast, "sensing (S type)" users pay more attention to practical and current things. They may not be very interested in participating in discussions of such abstract concepts, so their number is relatively small in the dataset.

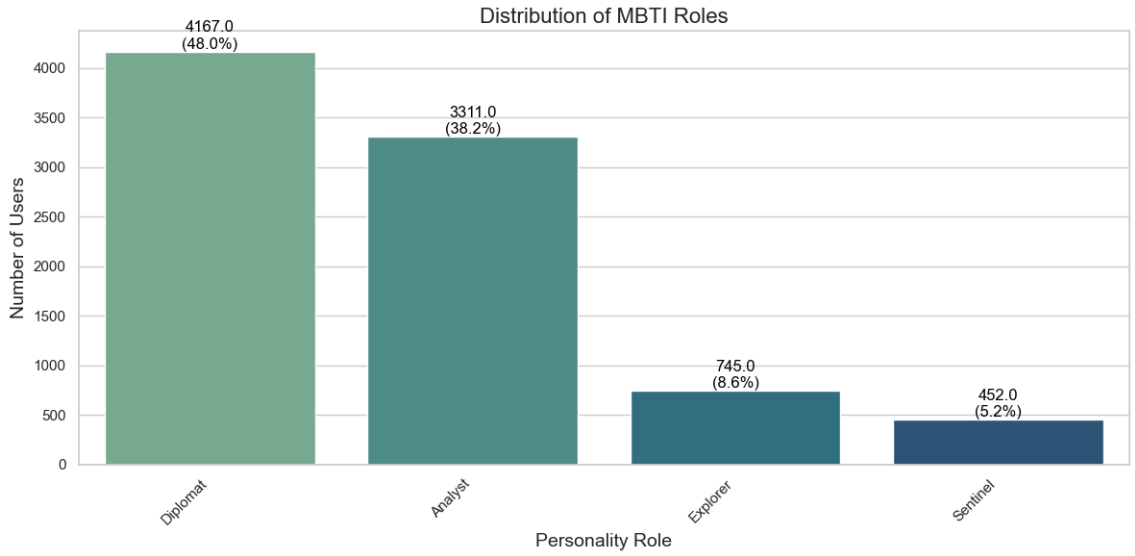

Figure 2.2: Distribution of MBTI Personality Roles

Based on the Figure 2.1, here's another bar chart (Figure 2.2) supplements the

proportion of users in the four main roles based on the 16 MBTI personalities, including Analysts, Diplomats, Sentinels, and Explorers. Among them, Diplomats accounted for the largest proportion, 48.0%, followed by Analysts, accounting for 38.2%. In contrast, Sentinels and Explorers accounted for a lower proportion, 5.2% and 8.6% respectively. This distribution further highlights the dominant proportion of "intuitive (N type)" users in the data set, which is consistent with the observations in the personality type distribution bar chart.

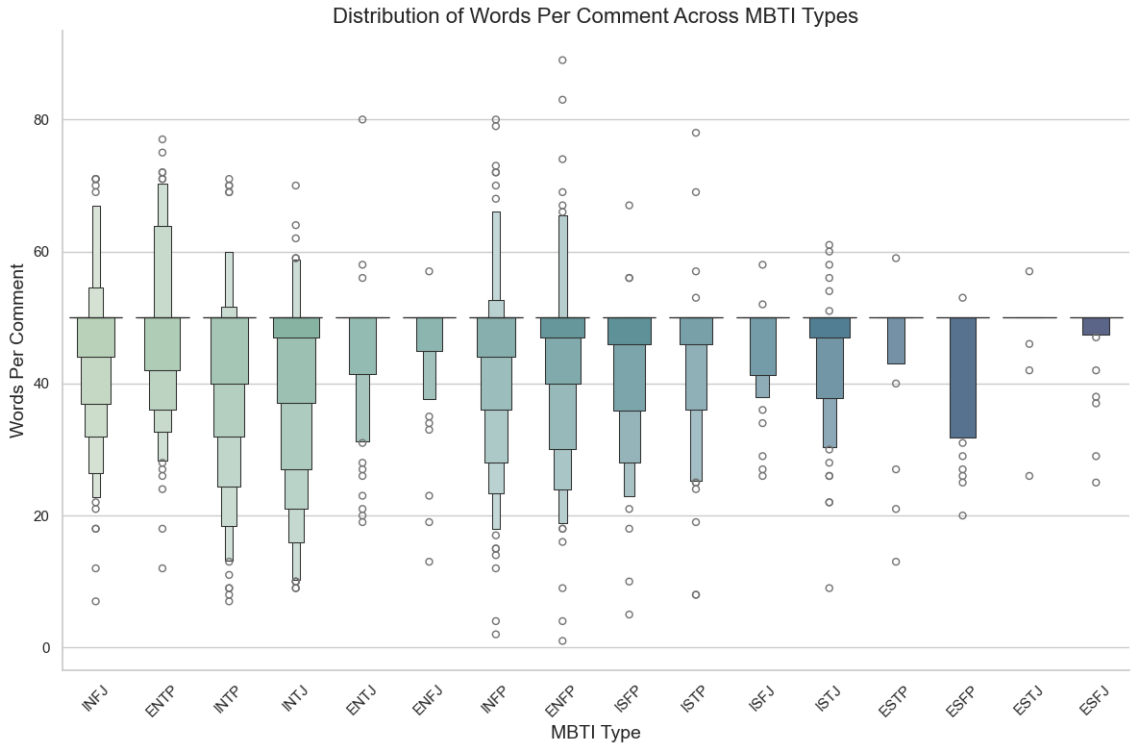

Figure 2.3: Distribution of Words Per Comment Across MBTI Types

The boxen plot (Figure 2.3) compares the word count distribution trend of each comment in the comments of users of 16 MBTI personalities. As can be seen from the figure, the median word count of comments of most personality types is concentrated

between 30 and 50 words. Among them, the word count distribution of INFJ and INTP shows obvious viriation. The "tentacles" of the box plot are longer and contain more outliers, indicating that these types of users tend to post longer comments. In contrast, personality types such as ISTJ and ESTP have a more compact distribution of comment word counts, indicating that these users' comments may be more concise and direct. The figure reveals the potential relationship between personality type and user expression. Intuitive (N-type) users tend to express complex or detailed views, while sensory (S-type) users tend to use concise language.

The following word cloud (Figure 2.4) shows the most common keywords in the comments of users of different MBTI types. Overall, words such as "think", "people" and "one" occupy a major position in all types, indicating that the topics discussed by users have certain commonalities. It might be due to the source of the dataset, which comes from a forum dedicated to discussing personality and psychological types. Therefore, the content of the discussion is highly thematic. Users in the forum often communicate around topics such as personality traits, which may be why some highly common words, such as "thinking" and "people", appear frequently in all types. Users with Feeling (F) personality types tend to use emotional words such as "love", "feel" and "friend", reflecting their emphasis on emotional connection in thinking and expression; in contrast, users with Thinking (T) personality types use relatively fewer emotional words.

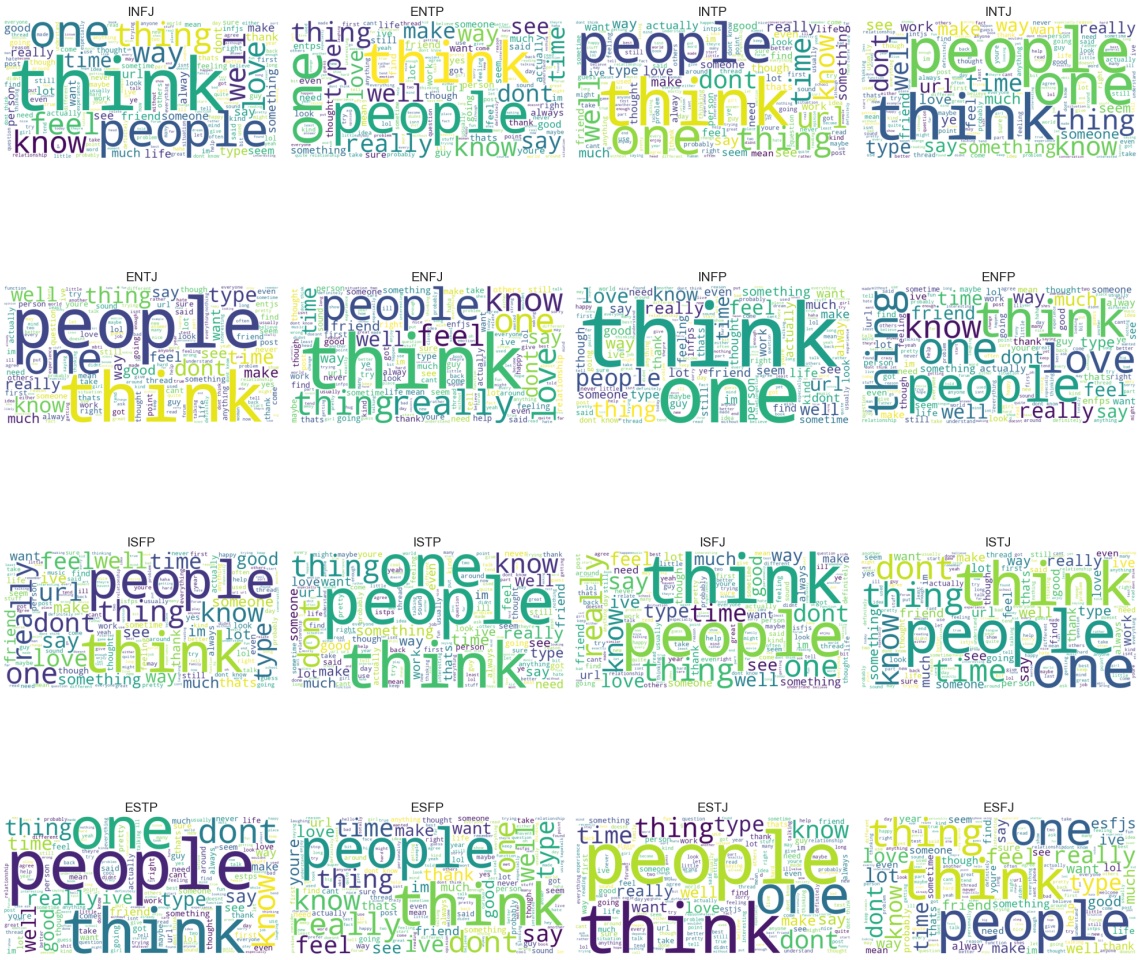

Figure 2.4: Most Frequent Words across Different MBTI Types

CHAPTER 3

Sentiment Analysis

After performing data preprocess and EDA, this chapter will introduce sentiment analysis in detail. Then, it will explain the implementation of sentiment analysis with MBTI dataset, and the results and findings.

Sentiment Analysis is a core task in the field of natural language processing (NLP), which can automatically identify and quantify emotions and attitudes in text, speech, or images. By analyzing text content, such as social media posts, sentiment analysis can help us understand people's emotion behind the text.

In sentiment analysis, model extracts feature from words and grammatical structures in the input text. Then, it uses specific models or pre-set dictionaries to determine the emotions behind the word. For example, "happy" and "happiness" often show positive emotions in the text, while "disappointment" and "pain" might related with negative emotions. The objective is to classify or score the text according to the emotional features contained. Therefore, we will get a more comprehensive understanding about the emotional expression tendencies reflected in human language.

3.1 Methods and Implementation

In this article, we will use the Myers-Briggs Type Indicator dataset to find the emotional tendencies across different MBTI types in social media. By performing sentiment analysis, we plan to find out which MBTI types of people are more likely to post positive comments, and which types of people are relatively more negative or neutral. This process can provide a useful reference for applications such as social science research and recommended algorithms.

In the process of implementing sentiment analysis, we used the VADER (Valence Aware Dictionary and sEntiment Reasoner) analyzer. VADER is a dictionary- and rule-based sentiment analysis tool designed for informal text environments such as social media. Its advantage is that it can be directly applied without additional training data, and it performs well when processing short texts with features such as emotional symbols, case deformation, and punctuation emphasis. Therefore, during the process of using the VADER analyzer for sentiment analysis, in order to maximize the advantages of the VADER analyzer, the dataset was further split. The 50 posts combined in the previous row were split into 50 rows. Therefore, the granularity of the dataset becomes higher, and the VADER analyzer can analyze the sentiment tendency of each independent post more carefully. VADER assesses the positivity and negativity of a single post more accurately, while avoiding the dilution or deviation of sentiment caused by too long text.

VADER calculates the overall sentiment score by assigning positive, neutral, and negative sentiment intensity values to each word and combining a series of rules. Finally, VADER outputs a normalized composite sentiment value (compound score) in the range of [-1, 1], where -1 represents very negative and 1 represents very positive.

The compound score calculation formula of VADER is:

$$\text{compound} = \frac{\sum_i \text{ valence }_i}{\sqrt{\sum_i (\text{ valence }_i)^2 + 15}} \tag{3.1}$$

3.2 Visualization Results and Analysis

After running the VADER analyzer, we get the returned sentiment score as the result, and group it by different MBTI types. Some of the results are posted to the table below. The table is made by 5 columns, "Clean_data" is the cleaned data after the original text data is cleaned and processed, and each row of data represents a single post, rather than 50. The remaining four columns are the negative sentiment, neutral sentiment, positive sentiment and compound score of the text.

cleaned_data	**neg**	**neu**	**pos**	**compound**
moment sportscenter top ten play prank	0.000	0.488	0.512	0.4939
lifechanging experience life?	0.000	1.000	0.000	0.0000
repeat today.	0.000	1.000	0.000	0.0000
may perc experience immerse you.	0.000	1.000	0.000	0.0000
last thing friend posted facebook ...	0.231	0.359	0.410	0.3612

Table 3.1: Sentiment Analysis(VADER) Results

We mainly use to the compound score of VADER and visualize it. After the above series of processing, we finally get the distribution diagram of VADER's compound score (Boxen Plot) and the compound score of VADER with different MBTI (Bar Plot)

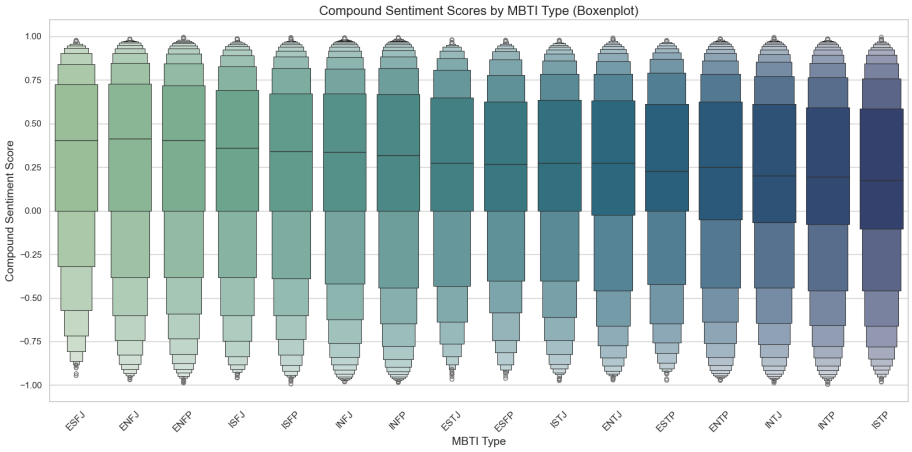

Figure 3.1: Compound Sentiment Scores by MBTI Types

The boxen plot (Figure 3.1) shows the distribution of VADER compound scores of various MBTI types in the form of boxenplot. The vertical axis is the sentiment score, which ranges from -1 to 1. The higher the score, the more positive the text content tends to be, and the lower the score, the more negative it tends to be. From the figure, we can see that for most MBTI types, the median of VADER compound score (the thicker position in the middle) is mostly close to 0 or slightly above 0, indicating that overall, most of the post sentiment tends to be slightly positive or close to neutral.

The score range of most types is relatively wide, ranging from extremely negative values close to -1 to extremely positive values close to +1. This means that people of all types may post relatively positive or negative content, but the thickness of the distribution and the concentration of the middle interval are slightly different. For example, the tail of the ESFJ boxenplot is relatively shorter and thinner, which means that the trend of ESFJ post and the distribution of VADER compound score are relatively concentrated.

The median range shows some subtle differences in tendencies. For example, the medians of ESFJ and ENFJ are in the middle of the 0.25-0.50 range, showing relatively high medians and a tendency to speak more positively. In contrast, the medians of INTJ, INTP, and ISTP are below the 0.25 range, showing relatively lower medians and a tendency to speak neutrally.

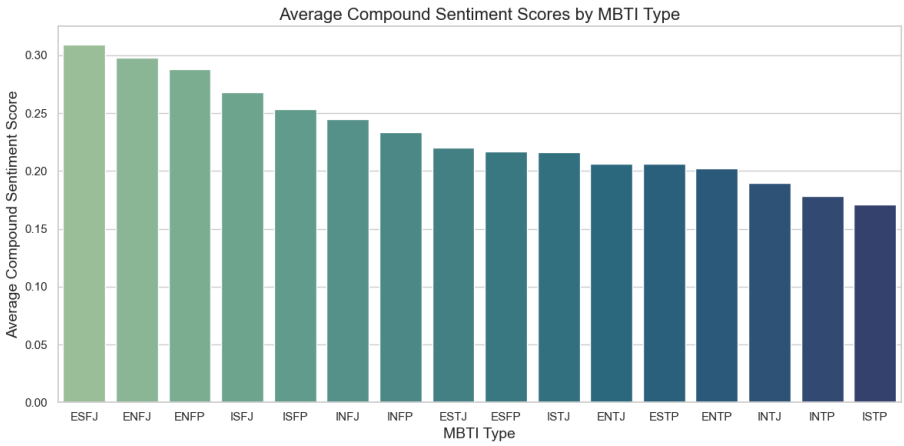

Figure 3.2: Average Compound Sentiment Scores by MBTI Types

The bar chart (Figure 3.2) shows the average comprehensive emotional scores of each MBTI type, and compares the average values in the form of a bar chart. This provides a more intuitive summary of the average tendency for the first distribution chart. The bar chart is sorted from left to right according to the score, showing a trend of average emotional scores from high to low. The ESFJ, ENFJ, and ENFP categories on the far left of the figure have the highest average scores, indicating that these types of users show a more obvious positive tendency in their overall posts. The INTJ, INTP, and ISTP on the far right of the figure have lower average emotional scores. Although they still have positive tendencies (scores greater than 0), these types of posts show a more neutral emotional tendency compared to the obviously

positive types on the far left.

Based on the above results and the visible overall trend, although it is impossible to draw an absolute emotional tendency conclusion through a single dimension (I/E, N/S, F/T, J/P), some related clues can still be seen from the performance of the average and distribution.

In the results, the more positive types are mostly concentrated in the types starting with E (such as ESFJ, ENFJ, ENFP), which shows that people with extroverted personalities tend to be relatively optimistic and positive in their expressions. Although there are some extroverted personality types whose average values are not particularly high, such as ENTP, overall, the E-type type is relatively high in the average emotional score, indicating that extroversion may be positively correlated with more positive language tendencies.

In addition, we can observe that among the types with higher VADER compound scores, the F (Feeling) personality is more prominent. Most of the top MBTI types (ESFJ, ENFJ, ENFP, ISFJ, ISFP, INFJ, INFP) contain F. This may be because the F type tends to be oriented towards emotions, empathy and harmonious interpersonal relationships, and is more likely to show positive, friendly and warm tendencies in language expression. Therefore, the F/T dimension seems to have a clearer correlation with emotional tendencies, and the F type generally presents a higher average positive emotional score.

From comprehensive observation, we can conclude that the types with high average positive emotion values are mostly E (Extrovert) + F (Feeling) personalities (such as ESFJ, ENFJ, ENFP) or at least F (Feeling) (such as ISFJ, ISFP, INFJ, INFP). This shows that people with a combination of extroversion (E) and feeling

(F) have more positive emotional comments on social media.

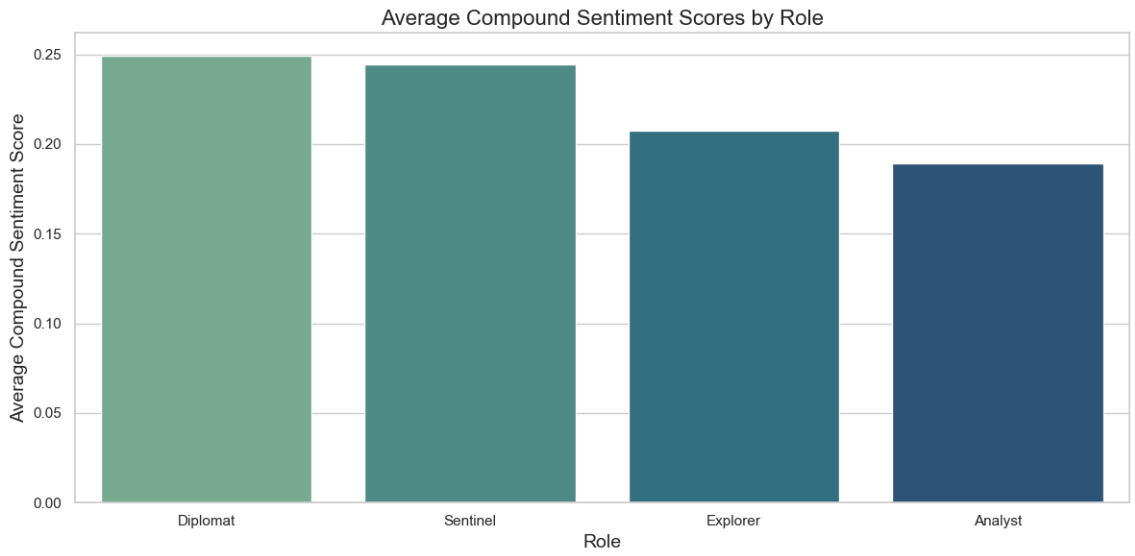

Figure 3.3: Average Compound Sentiment Scores by Roles

The figure 3.3 shows the performance of different personality roles in the Average Compound Sentiment Score, and compares them in the form of a bar chart. This visualization method allows us to intuitively compare the overall tendencies of each role type. From left to right, they are Diplomat, Sentinel, Explorer, and Analyst. It can be seen that the Diplomat type has the highest average score, close to 0.25, which shows that people of the Diplomat type tend to communicate in a more positive and optimistic language. The Sentinel type follows closely, with an average score slightly lower than the Diplomat, but still in a relatively high positive range, indicating that the Guardian type is still full of a certain positive tendency in the overall post.

The average sentiment score of the Explorer type has dropped slightly. Although it is still positive, compared with the obvious positivity of the Diplomat and Sentinel, the language tendency of this type of personality is closer to a neutral attitude rather

than a strong positive emotion. Finally, the Analyst type, although their average scores are all in the positive range, are significantly lower than the other three types. This means that the Analyst type is more neutral and calm in language expression, and lacks obvious positive preferences.

3.3 Integration of Overall Patterns and Insights

Comparing these trends with the previous segmentation type analysis, although it is not possible to conclude that a certain type is absolutely more positive from the four dimensions of E/I, N/S, F/T, and J/P, from the overall trend, Diplomat and Sentinel are more inclined to show positive, harmonious and positive verbal communication in the personality model, while Explorer and Analyst are relatively closer to neutrality and rationality. In this dimension, the emotional orientation or extroversion tendency in personality type may play a certain role, and the characteristics of feeling (F) and extroversion (E) are often positively correlated with more positive verbal expression.

In general, these results tell us that although a person's emotional tendency cannot be absolutely determined from a single dimension (such as I/E, N/S, F/T, J/P), the data still reveals some interesting clues. First, people who tend to be extroverted (E) appear to be more positive overall, whether ESFJ, ENFJ or ENFP, have higher average emotional scores. This may mean that extroverts tend to express themselves more in positive and optimistic language on social platforms.

At the same time, people of the "F" (Feeling) type also show a clear positive emotional tendency. For example, ESFJ, ENFJ, ENFP, ISFJ, ISFP, INFJ and INFP all have higher average emotional scores. This may be because they are more likely to empathize and care about other people's feelings, and naturally release more positive

and friendly atmosphere in their verbal expressions.

In other words, if a person is both extroverted and feeling-oriented (E+F), their verbal content tends to be more positive. Although not all extroverted types have particularly high scores (such as ENTP's scores are not so prominent), the overall trend still exists, and the influence of F traits seems to be more stable. In short, from this analysis, we can vaguely see that in the language expression of social platforms, extroverted and feeling-oriented traits tend to make people more positive, while introverted or rational thinking (T) tends to be more neutral and calm in expression style.

Among the four types of roles, Diplomat and Sentinel have higher average emotion scores, showing a more optimistic and positive language style; Explorer and Analyst have slightly lower average scores, and are more neutral overall. These results echo previous type analysis (such as the results of each subtype of MBTI), and once again provide clues about the correlation between personality tendencies and language emotion expression from a more macro level of role classification.

CHAPTER 4

Conclusion and Future Studies

This study analyzes the relationship between MBTI personality types and emotional expression on social media, revealing how personality traits affect language expression style. The study shows that people with extroversion (E) and emotional orientation (F) tend to express more positive emotions on social media, reflecting their more empathetic and optimistic personality traits. On the other hand, people with introversion (I) and rational thinking (T) tend to use neutral language, which is consistent with their more analytical and introverted personality.

In addition, through data distribution analysis, the study found that intuitive (N) personality users accounted for a high proportion in the data set, which may be related to the characteristics of the online forum where the data was sourced. As a forum dedicated to discussing personalities, Personalities Cafe attracts users who are more interested in abstract concepts and in-depth discussions. This coincides with the interests of intuitive (N) personality users.

Through VADER-based sentiment analysis. We can conclude that the combination of extroversion (E) and emotional orientation (F) is more likely to express positive emotions in language expression. For example, the sentiment scores of users of types such as ESFJ, ENFJ, and ENFP are significantly higher than those of other types. This tendency may be related to the social initiative of extraversion and the

emphasis on interpersonal relationships and higher empathy of emotion orientation.

In the MBTI role classification, the emotion scores of Diplomat and Sentinel roles are generally higher, showing a more positive and harmonious communication tendency. In comparison, the emotion scores of Explorer and Analyst roles are relatively low, and the language expression tends to be neutral and rational. However, the emotion scores of the four roles are all above 0, which means that everyone's social media post tends to be positive.

The findings of can provide practical application value in recommended algorithms and social media platform optimization. By analyzing the personality type and emotional tendency of users, social media platforms can more accurately predict users' preferences. Then, the recommended algorithm will be able to be strengthened and the accuracy could be increased. For example, for users with extroversion (E) and feeling (F) personalities, the content which is more on emotional communication may be more popular. For users with introversion (I) or thinking (T) personalities, the platform can push content that is more logical and in-depth. Then, the recommendation algorithm can more accurately push content that users are interested in. The user experience will be improved, thereby increasing the user's usage time. Furthermore, the user stickiness of social media platforms will also increase. Consequently, the great loyalty of users could build the foundation for the long-term development of the platform.

In marketing and user behavior analysis, the findings also has great potential. Brands and advertisers can develop more targeted marketing strategies based on the personality type and emotional tendencies of users. For example, the users who have the feeling (F) personality may be more easily moved by sensational advertisements.

In comparison, the users who have the thinking (T) personality tend to pay attention to advertising content with data support and clear logic. Thus, the targeted marketing can effectively improve conversion rate and brand loyalty.

In addition, in the field of mental health and counseling, sentiment analysis can help identify users who may have emotional distress on social media. Then, the text-detection algorithms can provide early warning for mental health intervention. At the same time, therapists can also use these technologies to analyze users' language expression and emotional state more efficiently. Thus, they can develop personalized counseling plans. The method of analyzing social media posts through sentiment analysis can provide strong data support for platforms and mental health services.

The future research can try to collect and utilize more diverse data sets, such as the post data from different social media platforms. In addition, future research can try to explore the interaction between personality traits and situational variables, such as the differences in emotional expression in different cultural backgrounds. Through these efforts, the research can more comprehensively reveal the correlation and influence of personality types on language and emotional expression. Finally, it can provide new research directions and application for the development of related fields.

REFERENCES

[16P] 16Personalities. "Roles Defined." https://www.16personalities.com/articles/roles-defined. Accessed: October 23, 2024.

[Dat] DataReportal. "Global Social Media Statistics." https://datareportal.com/social-media-users. Accessed: November 10, 2024.

[J17] Mitchell J. "(MBTI) Myers-Briggs Personality Type Dataset." https://www.kaggle.com/datasets/datasnaek/mbti-type, September 2017. Accessed: September 19, 2024.

[Per] My Personality. "Type Groups Theory." https://mypersonality.net/theory/type-groups. Accessed: October 23, 2024.