Data Analysis of the MBTI Personality Distribution of Telecommunication Fraud Victims

Kang chen*, Chuang Ma

Chongqing University of Posts and Telecommunications, Chongqing, China

*Correspondence author: joel6embiid@163.com

Abstract. In recent years, with the continuous development of technology, the methods of telecommunication fraud have increased, and the number of cases has shown a high incidence trend. By extracting the MBTI personality traits of victims and analyzing their personality types, it is possible to accurately prevent the personality traits of people who are susceptible to deception. Therefore, this paper uses the Bert model to extract the MBTI personality traits from the telecommunication fraud dataset to study how to fundamentally prevent the occurrence of telecommunication fraud cases. It analyzes the potential connection between MBTI personality traits and the group of victims of telecommunication fraud. Among more than 20,000 data points, INTP type victims were the most numerous, accounting for 34.5%. INFP types accounted for 27.8%, INTJ types for 13%, and INFJ types for 8.4%. Combined with the analysis of telecommunication fraud data from a specific area, there is a potential correlation between the probability of being defrauded, the methods of fraud, and personality traits in telecommunication fraud. Our research results may provide reference value for the government and relevant departments in managing telecommunication fraud.

Keywords: Telecommunications Fraud, MBTI Personality Traits, Bert Model, Fraud Prevention.

1 Introduction

As society continuously develops, fraudulent criminal cases emerge in an endless stream, with a wide variety of scamming methods. Among all types of crimes, telecommunication fraud stands out as the one with the highest occurrence and largest losses, with the involved amounts being difficult to recover. From 2019 to 2021, the prosecution authorities charged 39,000 people, 50,000 people, and 40,000 people respectively for telecommunication and internet fraud crimes. Although the number of people prosecuted in 2021 decreased, it still operated at a high level. Telecommunication and internet fraud involves many areas of daily life [1], with 50% of the scams being carried out under the guise of investment finance, emotional socializing, online shopping, etc., with investment finance scams ranking first, accounting for 26%.

With the government's intensified efforts to arrest perpetrators, in 2023, Chinese prosecution authorities charged more than 34,000 people for telecommunication and

Y. Kuang et al. (eds.), *Proceedings of the 2024 5th International Conference on Education, Knowledge and Information Management (ICEKIM 2024)*, Atlantis Highlights in Computer Sciences 22,
https://doi.org/10.2991/978-94-6463-502-7_81

internet fraud crimes, a nearly 52% increase compared to the previous year. Despite the increased intensity of the crackdown, the number of cases remains high. Therefore, analyzing the personality traits of victims of telecommunication fraud is crucial for preventing and combating fraudulent activities. First, it allows for an understanding of the characteristics of the susceptible group. By analyzing the personality traits of victims, it can be determined which personality types are more likely to be targeted by scams. This helps identify the most vulnerable groups, thereby making preventive measures more targeted. Secondly, it aids in customizing anti-fraud education and propaganda. Different personality traits mean different information processing and decision-making patterns. Understanding these traits helps in customizing effective education and propaganda strategies to enhance these groups' awareness of telecommunication fraud. Lastly, it contributes to improving the efficiency of fraud detection technology. Considering the personality traits of victims when developing automated tools and algorithms to identify potential fraudulent behavior can enhance the efficiency and accuracy of these tools.

The Myers-Briggs Type Indicator (MBTI) is a widely used psychological assessment tool designed to help people understand their own and others' personality types. It is based on the theories of Swiss psychologist Carl Jung and further developed by Katharine Cook Briggs and Isabel Briggs Myers. MBTI classifies individuals into 16 different types, each based on a combination of four binary dimensions: (1) Extraversion (E) versus Introversion (I); (2) Sensing (S) versus Intuition (N); (3) Thinking (T) versus Feeling (F); and (4) Judging (J) versus Perceiving (P). Y Li et al. [2] conducted an in-depth analysis in 2023 on the relationship between various personality theories and antisocial behavior. The article first notes the recent increase in crime rates and antisocial behaviors, emphasizing the role of different personality theories in understanding antisocial behaviors. These theories include the Myers-Briggs Type Indicator (MBTI) and the Big Five personality traits. The article explores how these personality characteristics influence antisocial behavior and suggests some feasible measures and solutions to mitigate antisocial behaviors. Daneshnari et al. [3] investigated the potential of the MBTI questionnaire in predicting criminal behaviors among students of Ferdowsi University of Mashhad. The study aimed to analyze the application of MBTI personality types in the context of Pinatel's theory of criminal personalities. Chuang Ma et al. [4] based their research on the Big Five personality theory, combining the Bert and SVM models to propose a model for identifying personality traits. They analyzed the association between the probability of being deceived, the methods of deception, the locations of telecommunication fraud, and personality traits.

With the ongoing development of machine learning technology, an increasing number of researchers are using these techniques to predict personality traits based on unstructured data such as text and video, applying them in fields like corporate human resources, education, and crime prevention. Mairesse observed [5] that individuals with high openness traits tend to use words related to insightfulness, while those with a tendency towards neuroticism are more likely to use more specific and common words in their writings. Neurotic traits are also associated with the use of words that carry negative evaluations and impacts. Automated personality detection methods leverage these manually inferred insights and related vocabulary as feature-based inputs for machine learning classifiers. For instance, linguistic inquiry and word count (LIWC) and the MRC psycholinguistic database dictionary are used for automatic predictions of social media texts [6].

Verhoeven et al. [7] introduced an integrated model for extracting personality traits from Facebook texts. The author trained three SVM classifiers, with the first being trained on Facebook. The second was trained on the Essays dataset as a meta-classifier, and the outputs of the two classifiers were fed into the third SVM classifier. Wright et al. [8] used part-of-speech and n-grams, negation, and vocabulary richness features to predict personality, employing the Big Five personality model. They collected 2588 essays from college students to evaluate the method, and the results showed that using these features could effectively improve prediction accuracy.

In 2012, Randall et al. [9] applied data mining and machine learning techniques to predict users' personality traits, especially the Big Five model characteristics, using only demographic and text-based attributes extracted from users' profiles. Darliamsyah et al. [10] developed an emotion-based personality detection system. Their findings indicate that individuals with the same personality traits tend to express emotions in similar ways, demonstrating the correlation between personality and emotions. Darliansyah and others proved that emotions, as a state of a person at a particular moment, are closely related to the long-term manifestations of personality traits. Therefore, emotional characteristics can be effectively used for the analysis in personality detection.

Many researchers have combined personality traits with crime prevention. Daneshnari et al. [11] conducted an MBTI test on undergraduate law students at a university through a questionnaire survey. They explored how individual personality traits influence the tendency towards criminal behavior, finding that individuals with feeling personalities are more likely to commit crimes due to emotionalism and low self-control. Conversely, individuals with intuitive and judging personalities avoid crime through rational decision-making, realism, and high self-control. These findings support the potential of using the MBTI personality as a tool for crime prediction and risk management, highlighting the significant role of individual personality traits in crime prediction and providing a basis for taking preventative measures. Li et al. [12] explored the association between personality traits and antisocial behavior, mentioning that there is a close connection between MBTI personality and antisocial behavior, especially certain personality types like ISTJ and INTJ, which may be more prone to displaying antisocial behavior due to a lack of empathy. Li emphasized that early detection and intervention in these personality traits and antisocial behavior connections could prevent the occurrence of such behaviors. These studies lay the foundation for applying the MBTI personality to telecommunication fraud in this paper. The main contributions of this paper are as follows:

(1) Using a fine-tuned BERT model to predict MBTI personality traits in a telecommunication fraud dataset.

(2) Analyzing a telecommunication fraud dataset from a specific area to uncover potential correlations between MBTI personality traits and case information, providing references for the prevention of telecommunication fraud.

The rest of this paper is organized as follows: Chapter two provides an overview of pre-trained models and the MBTI psychological model. Chapter three preprocesses the fraud dataset from a certain city. Chapter four predicts on the fraud dataset using BERT and analyzes the experimental results. Finally, it concludes the contributions of this paper to the prevention of telecommunication fraud, offering reference methods for related departments to prevent telecommunication fraud in the future.

2 Literature Review

2.1 A Review of the Literature on Pre-trained Language Models

In the field of Natural Language Processing (NLP), pre-trained models have become a significant research direction in recent years, especially with the advent of the BERT (Bidirectional Encoder Representations from Transformers) model, which has ushered in a new chapter for deep learning in NLP. Introduced by Google in 2018 and based on the Transformer architecture, BERT learns rich language representations through pre-training on a large corpus and then fine-tunes for downstream tasks, demonstrating exceptional performance [13]. The introduction of this model has not only led to leaps in performance for language understanding tasks but also spurred a series of BERT-based variants and related research.

The core idea of BERT is to capture the deep bidirectional contextual relations in text through a bidirectional Transformer encoder. Unlike previous models, BERT employs two tasks during its pre-training phase: the Masked Language Model (MLM) and Next Sentence Prediction (NSP). The MLM randomly masks part of the input sequence and asks the model to predict these masked words, effectively utilizing bidirectional context. The NSP task involves predicting whether two given sentences are in a sequential textual relationship, enhancing the model's ability to capture relationships between sentences. This pre-training mechanism enables BERT to learn rich language features and deep contextual dependencies, providing a robust foundational representation for various NLP tasks.

Since BERT's introduction, it has achieved unprecedented results in multiple NLP tasks, including but not limited to text classification [14], question answering systems [15], named entity recognition [16], and sentiment analysis [17]. In these tasks, BERT adjusts its pre-trained parameters during the fine-tuning phase to meet the specific needs of the task, demonstrating strong generalization capabilities and flexibility.

2.2 Literature Review of the MBTI Personality Model

Starting in 2023, the Myers-Briggs Type Indicator (MBTI) psychological model became a global phenomenon, with people eagerly testing their MBTI personality traits through various means. People began using their MBTI traits to make friends, seek jobs, and improve communication and interactions with others.

The Myers-Briggs Type Indicator (MBTI) is a widely used psychological model [18] that helps individuals understand personal differences and their impacts on work, interpersonal relationships, and personal growth. MBTI classifies individuals into 16 different personality types across four dimensions, each with its unique characteristics and tendencies. Here are the four dimensions of MBTI:

(1) Extraversion (E) versus Introversion (I): This dimension describes how individuals gain energy. Extraverts tend to draw energy from the external world and interactions with people, while introverts obtain energy from their inner world and solitude.

(2) Sensing (S) versus Intuition (N): This dimension concerns the processing of information. Sensing individuals focus more on the actual information obtained directly through the senses, while intuitive individuals rely on intuition and the interpretation of what information might mean.

(3) Thinking (T) versus Feeling (F): This dimension describes the decision-making process. Thinking individuals rely more on logic and objective analysis when making decisions, while feeling individuals consider interpersonal relations and emotional factors more.

(4) Judging (J) versus Perceiving (P): This dimension relates to attitudes towards the external world. Judging individuals prefer a planned, orderly way of life, while perceiving individuals are more flexible and open.

MBTI is extensively applied in areas such as personal development, team building, career planning, and management of interpersonal relationships. By understanding their own and others' MBTI types, individuals can better recognize their strengths and potential areas for growth. Within organizations and businesses, MBTI is often used in team-building activities to help team members understand each other's differences, fostering team cooperation and efficiency.

3 Data Preprocessing

The fraud dataset used in this article contains detailed information on fraud cases that occurred in a certain city from 2012 to 2020, with over 50,000 case records in total. Each record includes data such as the time and location of the fraud, the victim's ID, the fraudster's ID, and a description of the case. The dataset also includes some unsolved cases, for which the fraudster's ID is marked as missing and represented by a 0. The temporal and spatial distribution of the dataset is illustrated in Figures 1 and 2.

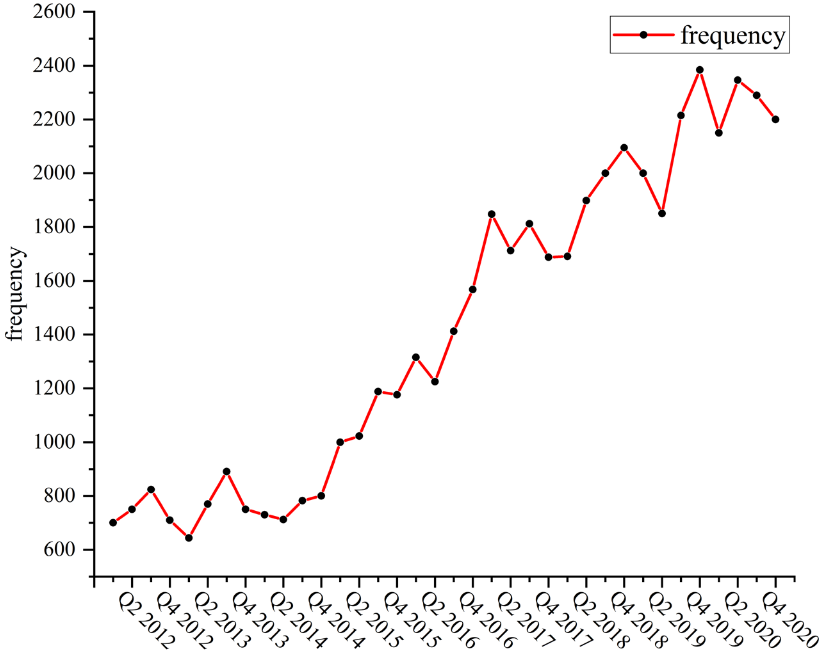

Fig. 1. Trends in the number of fraud cases in a place from 2012 to 2020

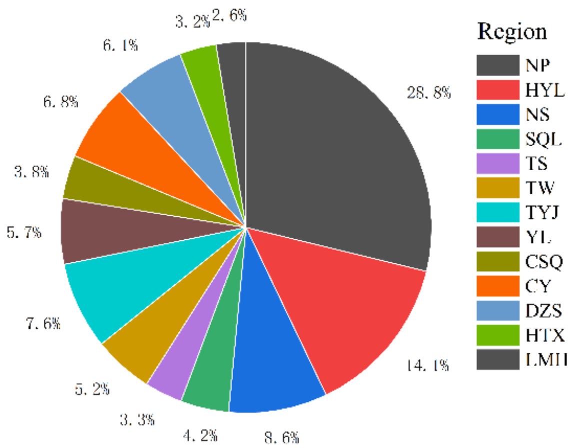

Fig. 2. The proportion of fraud cases in each period

This study primarily uses the detailed description text of cases provided by victims to predict their MBTI personality traits. Since the dataset was collected by public security departments, to ensure the privacy of the data, sensitive information such as bank card numbers, mobile phone numbers, and addresses were removed using regular expressions. Considering that each victim's description contains a lot of information irrelevant to predicting MBTI personality traits, to reduce the model's sensitivity to specific numbers and focus on the semantic content of the text, all numbers were replaced with the "[number]" token. Information that could lead to model overfitting, such as dates and specific place names, was excluded to improve the model's generalizability. Subsequently, the BERT pre-trained model's tokenizer was used to tokenize the text, transforming it into token sequences that the model can process. These tokens were further encoded into corresponding IDs, preparing them for model training. Given the BERT model's strict requirements on input length, this study standardized the length of all text sequences. Texts exceeding the predetermined length (e.g., 512 tokens) were truncated, while shorter texts were padded to ensure a consistent length for all input sequences.

4 Experimental Verification

The MBTI dataset on Kaggle typically includes text data collected from social media, forums, and other online platforms, with each piece of text annotated with the author's MBTI type. It contains 8,675 rows of data, where the first column is the MBTI type, and the second column contains individual posts (up to 50 items), separated by "|||". Due to the time-consuming nature of predicting over 50,000 data points, the experimental section of this paper selected more than 20,000 telecommunication fraud data entries from 2018 to 2020.

For the experiments, this paper employed the BERT BASE model, fine-tuning it on the labeled Kaggle MBTI dataset. Since the task is to predict 16 different MBTI personality traits, the fully connected output layer of the model was set to have 16 neurons. The trained model was then applied to our fraud dataset to predict MBTI types, thereby analyzing the MBTI personality types of telecommunication fraud victims and identifying potential correlations between MBTI personality traits and susceptibility to telecommunication fraud.

The analysis focused on identifying which MBTI personality types are most susceptible to fraud and observing trends over time. By examining the number of victims for each MBTI type, the research identified the types with the highest number of victims and observed their trends over time, analyzing which MBTI types may be more likely to become targets of telecommunication fraud. This approach aims to provide insights into the characteristics that make certain personality types more vulnerable to fraud, potentially guiding preventative measures and targeted awareness campaigns.

4.1 Victims' MBTI Personality Distribution

Based on the experimental results, we first tallied the total number of victims for each MBTI type to identify which personality types had the highest number of victims. Then, we analyzed the trends in these numbers over time. According to the analysis, from the first quarter of 2018 to the fourth quarter of 2020, the INTP type had the highest number of victims, totaling 8,670 cases. The INFP type followed closely behind with 6,968 cases. The INTJ type ranked third with 3,277 cases, and the INFJ type also had a relatively high number of victims, totaling 2,108 cases. The remaining ten personality types had significantly fewer numbers. These data indicate that individuals with a preference for Intuition (N), especially those with IN-type MBTI personality traits, represent a higher proportion of victims in telecommunication fraud. This suggests that individuals with Introverted Intuition (IN) tendencies are more susceptible to telecommunication fraud. The distribution of MBTI personality traits among the victims is illustrated in the following figure 3.

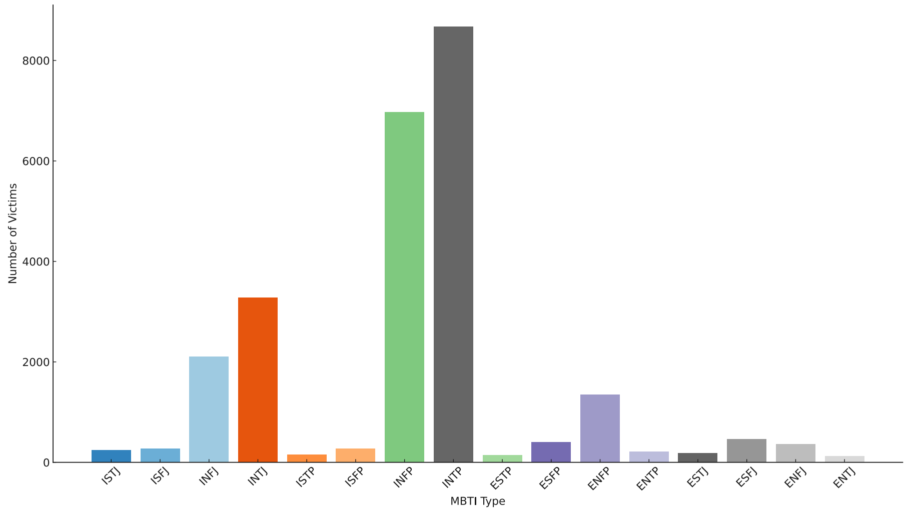

Fig. 3. MBTI personality distribution

Next, we analyzed the trends over time for the MBTI types most susceptible to fraud (INTP, INFP, INTJ, INFJ) to understand which personality types have seen the most significant increase in the number of victims as time progresses, as illustrated in Figure 4.

These trends suggest that individuals with Intuitive (N) and Introverted (I) MBTI personality traits seem to be more susceptible to telecommunication fraud. Particularly, individuals of the INTP and INFP types might be more vulnerable to fraudulent information or strategies due to specific personality traits such as high introspectiveness and the pursuit of ideals.

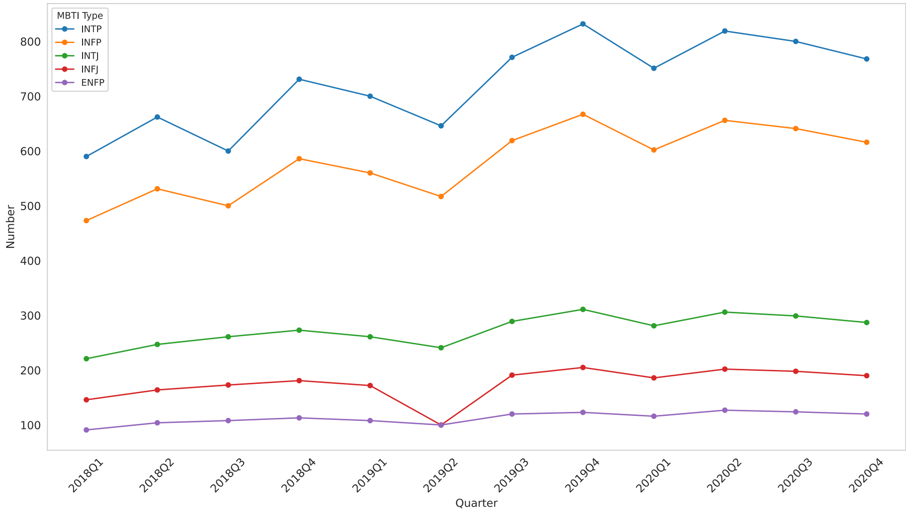

Fig. 4. MBTI personality distribution by quarter

In summary, the analysis of MBTI personality types among telecommunication fraud victims reveals that certain personality types—mainly those within the IN category—are more likely to become targets of scams. This finding can provide insights for the development of preventive measures, especially when designing awareness and educational materials targeted at these vulnerable groups. Over time, the trend of INTP and INFP individuals becoming victims has increased, which may reflect the evolution of telecommunication fraud tactics and the increased targeting of individuals with these specific personality types.

4.2 MBTI Distribution of Fraud Methods

As shown in the Figure 5, the INTP type has the highest proportion in the category of "gambling website" fraud methods, indicating that compared to other personality types, INTPs may be more susceptible to this kind of scam. INTPs are often considered to be curious, enjoy exploring, and solving problems, which might make them more interested in gambling or similar speculative ventures. INFPs, known as idealists, may be more easily deceived by scam schemes offering financial assistance. The distribution of various personalities across different fraud methods varies, suggesting that different personality types may have different sensitivities to various scam tactics. From this data, we can infer that devising targeted prevention strategies could be more effective for people of different MBTI types. For example, prevention strategies for INFP types might need to emphasize the risks of financial fraud and identification skills, while strategies for INTP types might focus more on explaining the probabilities involved in gambling and risk management.

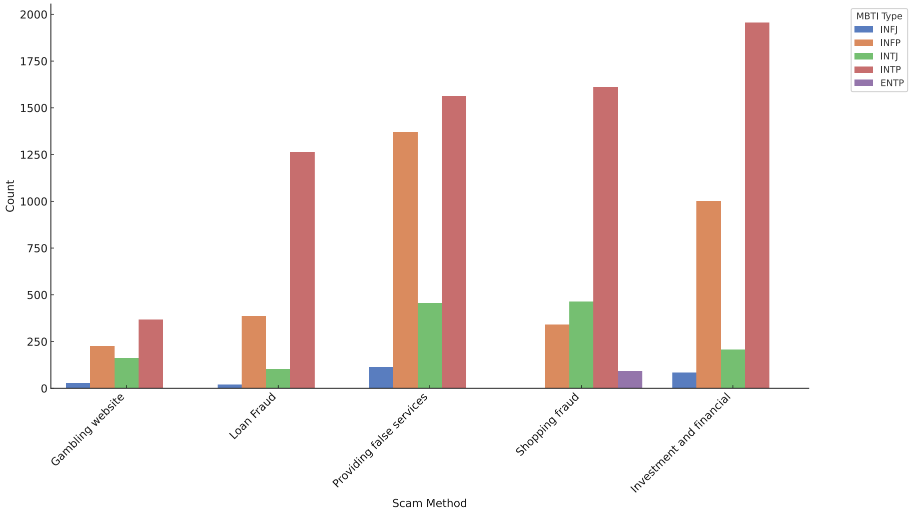

Fig. 5. MBTI personality types' susceptibility to different fraud methods.

Through these specific data and analyses, we can see that while anyone is susceptible to becoming a victim of fraud, different MBTI personality types may exhibit varying sensitivities to different types of fraud due to their unique character traits. This understanding helps individuals recognize their potential vulnerabilities and provides more targeted information for prevention education.

5 Conclusion And Future Work

This paper utilizes the Bert pre-trained model to extract the MBTI personality traits of victims from a telecommunication fraud dataset in a certain city. It then analyzes the distribution of the victims' MBTI personality numbers, with the most susceptible MBTI types being INTP, INFP, INTJ, and INFJ. Additionally, this paper examines the relationship between MBTI personality traits with time and fraud methods, providing references for more targeted preventive measures in the future. In future research, we will further explore the complex relationship between personality types and telecommunication fraud. For example, considering more demographic variables, such as age, educational background, and geographical location, to deepen the understanding of the characteristics of telecommunication fraud victims.

Reference

1. PEI Bingsen, LI Xin, WU Yue. Influence Evaluation of Telecom Fraud Case Types Based on ChatGPT[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(10): 2413-2425. 10.3778/j.issn.1673-9418.2306044

2. Li Y. Associations between Personalities and Antisocial Behavior[J]. Journal of Education, Humanities and Social Sciences, 2023, 22: 825-831. https://doi.org/10.54097/ehss.v22i.14243
3. Daneshnari H, Sadeqnexad M, Marvi M. Crime Prediction Using the Myers-Briggs Personality Test Case Study of Ferdowsi University of Mashhad Students[J]. Criminal Law Research, 2023, 14(1): 63-88. 10.22124/JOL.2023.22164.2283
4. Gao J, Ma C, Hu J. Research and Analysis of Potential Correlation of Telecommunications Fraud Data Based on Big Five Personalities[J]. 2023. https://doi.org/10.21203/rs.3.rs-2392379/v1
5. Mairesse F, Walker M A, Mehl M R, et al. Using linguistic cues for the automatic recognition of personality in conversation and text[J]. Journal of artificial intelligence research, 2007, 30: 457-500. https://doi.org/10.1613/jair.2349
6. Tausczik Y R, Pennebaker J W. The psychological meaning of words: LIWC and computerized text analysis methods[J]. Journal of language and social psychology, 2010, 29(1): 24-54. https://doi.org/10.1177/0261927X09351676
7. Verhoeven B, Daelemans W, De Smedt T. Ensemble methods for personality recognition[C]. Proceedings of the International AAAI Conference on Web and Social Media, Cambridge, Massachusetts USA, 2013, 7(2): 35-38. https://doi.org/10.1609/icwsm.v7i2.14465
8. Wright W R, Chin D N. Personality profiling from text: introducing part-of-speech N-grams[C]. User Modeling, Adaptation, and Personalization: 22nd International Conference, Aalborg, Denmark, 2014: 243-253. https://link.springer.com/chapter/10.1007/978-3-319-08786-3_21
9. Poria S, Gelbukh A, Agarwal B, et al. Common sense knowledge based personality recognition from text[C]. Advances in Soft Computing and Its Applications: 12th Mexican International Conference on Artificial Intelligence, Mexico City, 2013: 484-496. http://dx.doi.org/10.1007/978-3-642-45111-9_46
10. Darliansyah A, Naeem M A, Mirza F, et al. SENTIPEDE: A Smart System for Sentiment-based Personality Detection from Short Texts[J]. J. Univers. Comput. Sci., 2019, 25(10): 1323-1352. https://doi.org/10.3217/jucs-025-10-1323
11. H, Sadeqnexad M, Marvi M. Crime Prediction Using the Myers-Briggs Personality Test Case Study of Ferdowsi University of Mashhad Students[J]. Criminal Law Research, 2023, 14(1): 63-88. 10.22124/JOL.2023.22164.2283
12. Li Y. Associations between Personalities and Antisocial Behavior[J]. Journal of Education, Humanities and Social Sciences, 2023, 22: 825-831. https://doi.org/10.54097/ehss.v22i.14243
13. Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018. https://doi.org/10.48550/arXiv.1810.04805
14. Prabhu S, Mohamed M, Misra H. Multi-class text classification using BERT-based active learning[J]. arXiv preprint arXiv:2104.14289, 2021. https://doi.org/10.48550/arXiv.2104.14289
15. Rmnath S, Nema P, Sahni D, et al. Towards interpreting BERT for reading comprehension based QA[J]. arXiv preprint arXiv:2010.08983, 2020. https://doi.org/10.48550/arXiv.2010.08983
16. Liu Z, Jiang F, Hu Y, et al. NER-BERT: a pre-trained model for low-resource entity tagging[J]. arXiv preprint arXiv:2112.00405, 2021.

17. Acheampong F A, Nunoo-Mensah H, Chen W. Transformer models for text-based emotion detection: a review of BERT-based approaches[J]. Artificial Intelligence Review, 2021, 54(8): 5789-5829. 10.1007/s10462-021-09958-2
18. Boyle G J. Myers-Briggs type indicator (MBTI): some psychometric limitations[J]. Australian Psychologist, 1995, 30(1): 71-74.