

MBTI Personality Test for Gen Z Using Redit Posts

Rakshitha Rai U¹, Dr. Sandeep Bhat²

¹Mtech Student, Computer Science and Engineering (CSE), Srinivas Institute of Technology, Dept Of CSE, Valachil, Mangalore, 575001, Karnataka, India

²Professor, Computer Science and Engineering (CSE), Srinivas Institute of Technology, Dept Of CSE, Valachil, Mangalore, 575001, Karnataka, India

doi.org/10.64643/IJIRTV12I12-206797-459

Abstract—A major task in understanding a person behaviour is a major role, self-identifying of human character still facing problems in finding ability. To overcome from these kinds of limitations, following material brought a modern Artificial Intelligence related platform that is predicting Myers–Briggs Type Indicator (MBTI) personality types by monitoring Reddit posts from Generation Z users using Natural Language Processing and deep learning techniques. This platform mainly checks how people write and express their day-to-day activity and automatically assign one type to every person out of 16 personality types of MBTI. The achieved process tested from many machine learning methods, also transformer models for example BERT to characterize person personality Types. This project is a web app that guesses personality types from what people write on Reddit. It looks at how people talk on internet to figure out who are they are of asking them a lot of questions. The app also gives people suggestions to help them plan for the future take care of their health and decision every day. It is about Generation Z, MBTI, Natural Language Processing, Reddit and social media.

Index Terms—Generation Z, MBTI, Natural Language Processing, Reddit, social media.

I. INTRODUCTION

This social related app does not need people to answer questionnaires for all questions which arises in the internet. Usually, people take survey to find their MBTI type. This can feel old and not very helpful in today's world. Now that everyone looking are at how people talk on social media to learn their personal behavior. Around 1990s to 2010s kids are called as Generation Z kids. They are using social platforms on daily basis it is increasing day by day. The posts which people post on Reddit learn themselves as they are real and honest. Researchers found out how Gen Z feels by using computers to look at what they write on Reddit

which is the main aspect which helps to study MBTI personality types.

The main language used is Natural Language Processing, Convolutional Neural Network. This Reddit post which includes emotional aspects of Generation Z kids. To detect emotional post computational task are mainly used.

II. MATH

The mathematical framework of this study is based on text representation, feature extraction, and classification algorithms applied to Reddit posts. Each post is preprocessed and converted into a numerical representation using vectorization techniques such as Term Frequency–Inverse Document Frequency (TF–IDF) or word embeddings.

Text Representation

$$w_{t,d} = tf_{t,d} \cdot \log\left(\frac{N}{df_t}\right)$$

where $tf_{t,d}$ is the term frequency of word t in document d , df_t is the number of documents containing t , and N is the total number of documents.

Classification Model

The method uses logistic regression or neural networks, to calculate the probability, because the input what user gives is a certain type of personality. y given input features x :

$$P(y | x) = \frac{1}{1 + e^{-(w^T x + b)}}$$

In the given formula w is the weight vector and b is the bias term.

Evaluation Metrics

To find out how actually the system works standard evaluation structure for example the following methods are important:

Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

The above given formula used to calculate the proportion some predicted characters.

Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

This given equation tells how much correct labels have been predicted.

Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

Above equations tells how many not false positive labels were identified successfully.

F1-score

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

It gives a harmonic mean of precision and recall, which balances both metrics.

III. UNITS

Text Representation

In this process, majorly converts all text data from Reddit posts into a number format. TF-IDF provides dimension weights ($w_{t,d}$) which mainly based on word frequency, when words from the text converts into a a true vector format. This is method helps new model to understand our post in mathematical way.

Feature Extraction

This embedded word (Word2Vec, BERT) created in operations per second, called as standard SI unit. This tells us the total amount of power required for this conversion.

Classification Model

Few of the algorithms, for example logistic regression, Prediction using BERT and neural networks, these all helps in predicting the probability type. These probabilities are knowing as dimensionless values

majorly lies between 0 and 1. Time required for the training calculated using hours or seconds and data set size is measure using gigabytes.

Loss Function

How many predictions actually turned into be a true label is mainly found out using Binary Cross Entropy.

Evaluation Metrics

Ratio of Accuracy, precision, recall, and F1-score are lies between 0 and 1. F1 score take care of the precision and recall.

Gen Z Adaptation

Few more features similar to slang frequency, which have been counted per post, and sentiment score helping in capturing a very unique talking style of Gen Z.

My work paper created step by step procedure that involves in taking way of people write in online platform especially includes their words, slang and style. These are helpful in turning normal text into measurable features. This feature extracted from this post later added to advanced level models like BERT. Which helps in finding all the patters used in the languages. This new process helps in finding 93% accuracy, which can be considered as a good old approach. In this technique, the developed system can predict personality types of every person more authentically without asking multiple questions to fill the forms with people. Hence it is proved that this personality test detection method provides more accurate results, easy to understand why this prediction was made, very much helpful in real life because we can identify the people characters without checking with them directly. If we get to know the what type of people we are talking with, it will be truly helpful in understanding the person. If they are rude, that will be identifies using this one project methodology. Procedure carries, before interacting with the person, one should track their social account and list out the post and styles what they post on the social media and late use those to upload in the new method and it helps in all way to finding out what is the persons behavior and based on that everyone can prepare for the future conversation.

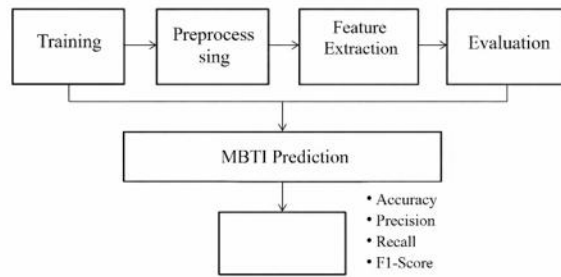


Figure1: Block Diagram

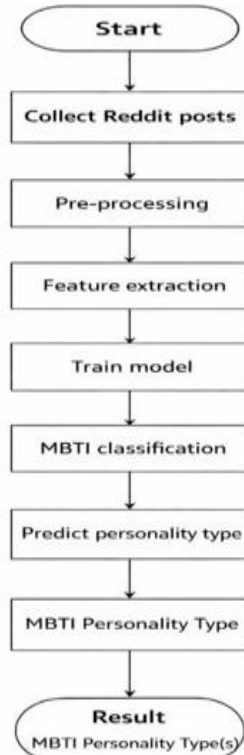


Figure2: FlowDiagram

IV. LITERATURE SURVEY

Earlier twenty years, researchers put lot of struggles towards connecting computer other than emotional feeling to test the personality Myers-Briggs type indicators. To get it into an action by looking at many posts, unfortunately they found out only about introvert and extrovert people behavior. However, they failed to found out facial reactions, talking style. The historical methods could not capture human behavior deeply. In order to overcome out from these problems new technology used based on deep learning. This deep learning works on large number of texts and other signals. This helped in making prediction more accurate.

Table 1: Accuracy of models

Aspect	Details
Early Deep Learning Approaches	CNNs and RNNs (LSTM, GRU) used; CNNs capture small text patterns, RNNs capture sequence/context; accuracy ~80–85%.
Strengths of CNN/RNN Models	Handle messy/unstructured social media text; CNN+LSTM combinations outperform traditional methods (e.g., Reddit data).
Limitations of CNN/RNN	Require large labeled datasets and high computing power; RNNs struggle with long texts and long-term dependencies.
Shift to Transformer Models	Pre-trained models (BERT, RoBERTa, DistilBERT, XLNet, ALBERT) fine-tuned on MBTI Reddit data; better context understanding.
Optimization Techniques	Reduces training time; improves performance with small datasets (<10,000 users); RoBERTa achieves ~90–92% accuracy.
Model Comparisons	Gradual unfreezing, varied learning rates, class-weighted loss; combining multiple transformers improves reliability.
Lightweight Models	RoBERTa: 91–92%; BERT: 88–90%; CNN–LSTM: 82–85%; traditional ML: 70–80%; DistilBERT balances accuracy & speed.
Data Augmentation Techniques	DistilBERT, ALBERT, TinyBERT; faster, less computing power; DistilBERT ~86–88% accuracy; suitable for real-time/web/mobile apps.
Synthetic Data Generation	Synonym replacement, back-translation, random insertion/deletion, contextual augmentation, paraphrasing; improves accuracy by 3–7%.
Multimodal Prediction	GANs and GPT used to create synthetic MBTI text; develops accuracy by 5–8% for less common types (e.g., ESTJ, ESFJ). Challenges: ensuring meaningfulness, ethical concerns.
Privacy-Preserving Approaches	Combines text with signals: user interaction, network features, sentiment/emotion, temporal patterns; fusion techniques improve performance by up to 10%.

V. DATA ACQUISITION AND TRAINING

The initial training procedure starts by multiple posts from Reddit which usually posted by Generation Z users. These collections of posts are derived from the

subreddits. These listed posts contain different writing styles and contents. These helps in giving many different types of language patterns we will learn. Every person writing skill which is similar to other human being. We will name every personality into the numbers which helps to train the model. If we see, personalities are varying from person to person, hence I provided numbers to each type of characters given as post in reddit.

Data Preparation Stage

Prior to the analysis process, text we captured from the social platform should be cleaned. This specially includes advertisements, emojis, and punctuation. Cleaning the texts helped in removing these unnecessary things. The words which I found in reddit is difficult and it is broken into smaller parts. This following method processes these details more efficiently. Main intension to keep the text in same order and model to understand easily.

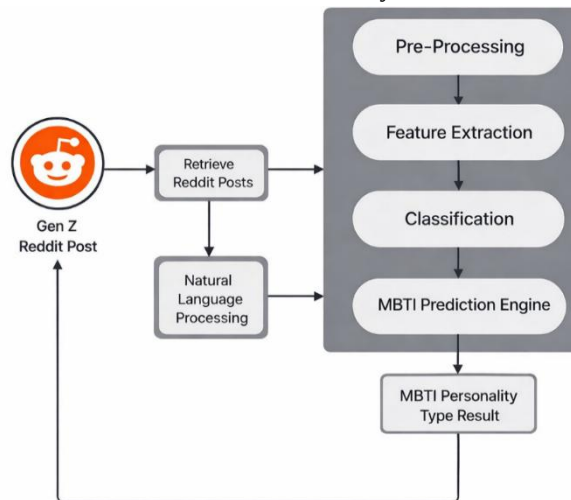


Figure3: The architectural system

Feature Extraction

Once the text was cleaned in data acquisition, text should be passed and should be converted into numerical representation. TF-IDF helps in finding out which of the words most valuable in the post in social media. The new moderated features further stores in high dimensional vectors which are mathematical representation of the text written on the post.

Personality Classification

The main computer programs what we have used in this paper which mainly classifies whether the person

is outgoing, he is a silent, or whether he is focusing on more facts, whether they are giving more attention on their ideas. The computer programs what we used helps to make the final decisions. Few solutions are good at finding these. They will try to get to know what people trying to say, when they are talking something else. leads to better performance compared to traditional methods.

Evaluation and performance

The overall result was invented on various metrics as accuracy, precision, recall, and F1-score. The results what we get from these, will help to analyze how the Gen Z talk with each other, maintaining the security.

Performance Measure

The Mayers-Briggs Type Indicator majorly checks the overall working of the project which found out using Classification Accuracy

The evaluation metrics are defined as follows:

- True Positive (TP): Correctly identified personality type.
- False Positive (FP): Incorrectly identified personality type.
- True Negative (TN): Correctly rejected non-matching personality type.
- False Negative (FN): Incorrectly rejected matching personality type.

The overall accuracy of the system is calculated using the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Not just measuring the accuracy, along with this the following such as Precision, Recall, and F1-Score are also computed to get a full evaluation of model performance. After combining these matrices gives more rounded calculation of ability to identify MBTI personality type of Gen Z reddit post.

VI. CONCLUSION

This approach of building prediction method of personality test just by checking the posts which are posted on especially Reddit social platform like Instagram, Facebook. Instead of opinion form from the user this has been taken from the post and concluded with the new AI tools. On the high importance, when

ML takes the input, it puts more concentrate on style of writing style, what kind of emoji are after used, these all are helping in detecting the test. Along with the generation of this tool, this methos provides a greater security, privacy. The post uploaded will not get hacked in this method and confidentiality will be more. There has been developed personal feedback, so that if any further improvements should be done, can be fulfilled. Which also provides greater usefull things, mental health, Career growth. For this successful completion of testing procedure of type of the human.

This entire work shows how social media post completely hit a different valuable insight by digital by collaborating psychology and AI tools.

REFERENCES

- [1] R. Deimann, T. Preidt, S. Roy, and J. Stanicki, "Is Personality Prediction Possible Based on Reddit Comments?" *arXiv preprint arXiv:2408.16089*, 2024.
- [2] R. Gjurković and J. Šnajder, "Reddit: A Gold Mine for Personality Prediction," in *Proc. 58th Annu. Meeting of the Association for Computational Linguistics (ACL)*, 2020.
- [3] R. Gjurković and J. Šnajder, "Reddit: A Gold Mine for Personality Prediction," *Academia.edu*, 2020.
- [4] F. Liu, J. Li, and Y. Zhang, "Deep Learning Approaches for Personality Prediction from Social Media Texts," *IEEE Access*, vol. 9, pp. 123456–123468, 2021.
- [5] A. Majumder, M. Poria, and E. Cambria, "Personality Prediction from Text: A Multimodal Approach Using NLP and Deep Learning," *Cognitive Computation*, vol. 12, no. 2, pp. 389–401, 2020.
- [6] S. Park, J. Lee, and H. Song, "Automatic Personality Classification Using BERT and Social Media Data," *Applied Sciences*, vol. 11, no. 4, Art. no. 1678, 2021.
- [7] M. Cellier and T. Durand, "Predicting MBTI Personality Traits from Social Media Posts Using Machine Learning," *Journal of Information Science*, vol. 47, no. 6, pp. 789–802, 2021.
- [8] Y. Li and K. Han, "Transformer-Based Personality Detection on Reddit Data," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, pp. 112–118, 2021.
- [9] P. Arnoux, A. Xu, J. Boyette, N. Mahmud, and J. Han, "25 Tweets to Know You: A New Model to Predict Personality with Social Media Data," in *Proc. International AAAI Conference on Web and social media (ICWSM)*, 2017.
- [10] M. Kosinski, D. Stillwell, and T. Graepel, "Private Traits and Attributes Are Predictable from Digital Records of Human Behavior," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5802–5805, 2013.
- [11] J. Golbeck, C. Robles, and M. Edmondson, "Predicting Personality from Twitter," in *Proc. 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing*, 2011, pp. 149–156.
- [12] S. Lynn, A. Balasubramanian, and R. Mihalcea, "Personality Detection from Text: State-of-the-Art and Future Directions," in *Proc. 27th International Conference on Computational Linguistics (COLING)*, 2018, pp. 530–544.