WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Personality Prediction using Social Media Data

Mrs. A. Sarala Devi, S. Vikas, P. Dhanush, Ch. Valli Srujana, P. Eashwar, V. Vamshi Reddy

Associate Professor, Department of CSE (Data Science), ACE Engineering College, Hyderabad, India

ABSTRACT

In the modern digital age, social media platforms have become significant repositories of behavioral data. The content shared through posts, captions, comments, and other interactions provides valuable insights into users' emotions, thoughts, and overall personality traits. These digital footprints can be utilized to understand individuals more deeply without the need for intrusive questioning or traditional psychological assessments.

This project focuses on developing a system that predicts user personality types using Natural Language Processing (NLP) and Machine Learning (ML) techniques. By analyzing textual data gathered from social media activity, the system maps linguistic features to established psychological models, specifically the Big Five Personality Traits—Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

The methodology involves data preprocessing, feature extraction from user-generated text, and classification using pre-trained ML models. Tools such as label encoders and serialized model files (e.g., Pickle) are used to ensure accurate and efficient personality prediction. This allows the system to provide quick and meaningful insights based on simple text input.

1. INTRODUCTION

This project aims to build an efficient **Personality Prediction System** using social media text data by applying **Natural Language Processing (NLP)** and **Machine Learning (ML)** techniques. The goal is to classify users based on the **Big Five Personality Traits**—Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

Key objectives include:

- Automatically analyzing user-generated text (posts, captions, comments) to predict personality.
- Providing **real-time predictions** with **confidence scores** for better reliability.
- Ensuring the system is **scalable** and can be integrated into fields like education, recruitment, and mental health.
- Designing a **user-friendly interface** for easy accessibility and interaction.

This system offers a non-intrusive, data-driven approach to understanding personality through everyday digital behavior.

2.EXISTING SYSTEM

- Traditional personality assessment techniques primarily rely on **self-report questionnaires**, such as the **Myers-Briggs Type Indicator (MBTI)** or the **Big Five Inventory (BFI)**. These assessments ask users to respond to a series of predefined questions that are analyzed to classify their personality traits. While widely used in psychology, education, and HR, these methods present several limitations in today's fast-paced, digitally driven world.
- Despite their usefulness, existing personality prediction systems face several limitations. Traditional methods rely heavily on self-report questionnaires, which are time-consuming, intrusive, and often suffer from social desirability bias—where users give answers they believe are more socially acceptable rather than truthful. These systems typically produce static results, failing to adapt to real-time changes in user behaviour or evolving personality traits. Additionally, they lack contextual understanding, often misinterpreting sarcasm, mood variations, or cultural nuances. Most platforms also provide limited interactivity and personalization, delivering plain reports without feedback or engagement. As a result, current systems fall short of delivering holistic, dynamic, and context-aware personality assessments.

3.PROPOSED SYSTEM

The proposed system is designed to predict users' personality traits and simultaneously assess their emotional well-being and stress levels using social media data. It leverages Natural Language Processing (NLP) and machine learning techniques to analyze the emotional tone of posts, comments, and language patterns related to stress and anxiety. Unlike traditional models that provide static personality profiles, this system offers a more holistic and dynamic psychological analysis. It is built with privacy in mind, incorporating user consent and data anonymization to ensure ethical usage.

Key Features:

This system combines personality prediction with emotional health monitoring, offering a dual-layered analysis. It detects stress indicators through text sentiment and behavioral cues, providing timely wellness insights. Key features include real-time emotional tone detection, mental health support recommendations, and privacy safeguards. The system is non-intrusive, scalable, and context-aware, making it suitable for applications in healthcare, education, recruitment, and personalized user experiences.

Advantages:

The proposed system enhances adaptability and usability in fields such as mental health, recruitment, personalized marketing, and user experience design. By analyzing social media behavior, it provides insights into both personality traits and emotional well-being. The system is designed to be intuitive and user-friendly, allowing for seamless integration into real-world applications where understanding user psychology is crucial for improving services and interactions.

4.METHODOLOGY

The Personality Prediction system follows a data-driven machine learning methodology. The approach involves analyzing user-generated textual data to classify personality types based on the Myers-Briggs Type Indicator (MBTI) framework.

The key steps in the methodology include:

Data Collection: The system uses an MBTI dataset (mbti_1.csv) containing textual posts and their corresponding personality labels.

Preprocessing: Textual data undergoes cleaning, tokenization, and vectorization using the TF-IDF (Term Frequency–Inverse Document Frequency) method to convert text into numerical feature vectors.

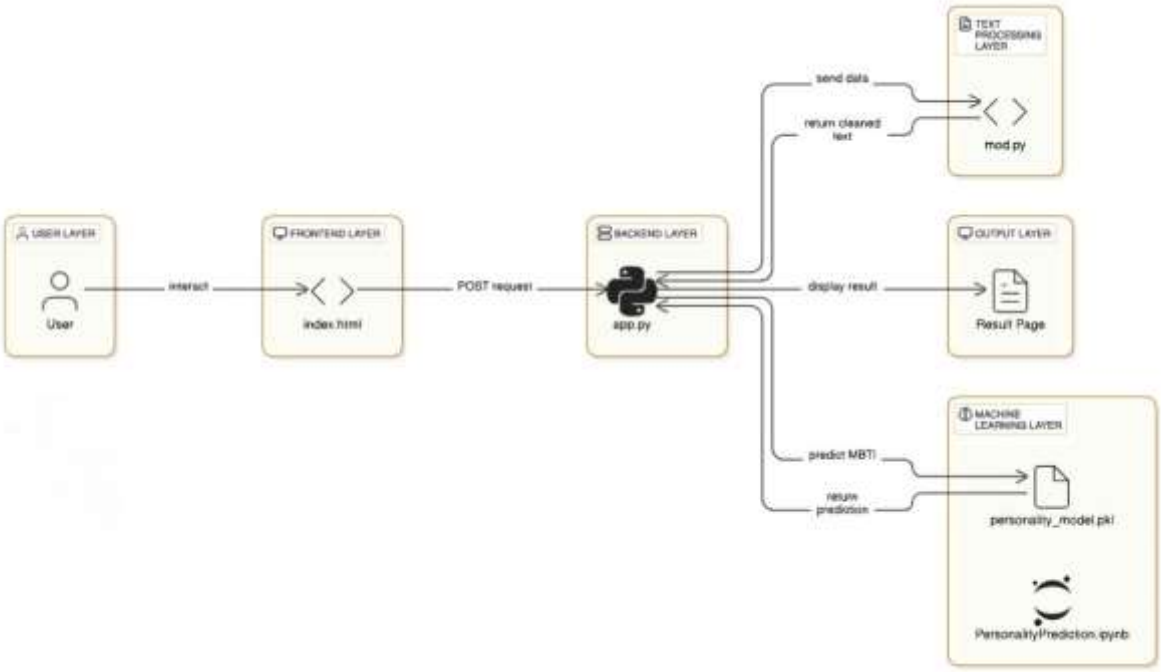
Model Selection and Training: A suitable classification algorithm (likely Logistic Regression, Random Forest, or SVM based on your model files) is trained on the processed data to learn patterns associating text features with personality types.

Model Serialization: The trained model and vectorizer are serialized using pickle for efficient storage and reuse in the web application.

Web Application Integration: A Flask-based web application (app.py) is built to provide a user-friendly interface for interacting with the prediction system. Users can input text data, which is processed and passed through the trained model to predict personality types.

5.SYSTEM ARCHITECTURE

System architecture is a comprehensive blueprint that defines the structure, behavior, and interactions of various components within a system—whether it's a software application, a computer system, or a complex network of systems. It provides a high-level view of how the system is organized and how different parts such as hardware, software, data storage, processing units, communication protocols, and user interfaces interact to perform specific functions. In software systems, architecture describes how modules or services are divided, how they communicate (e.g., via APIs or message queues), and how data flows through the system.



In hardware systems, it includes the design of processors, memory units, input/output devices, and how they are connected. System architecture also includes considerations for scalability (handling growth in users or data), security (protecting data and operations), maintainability (ease of updates and debugging), and performance (speed and efficiency).

6. RESULTS AND OUTPUT

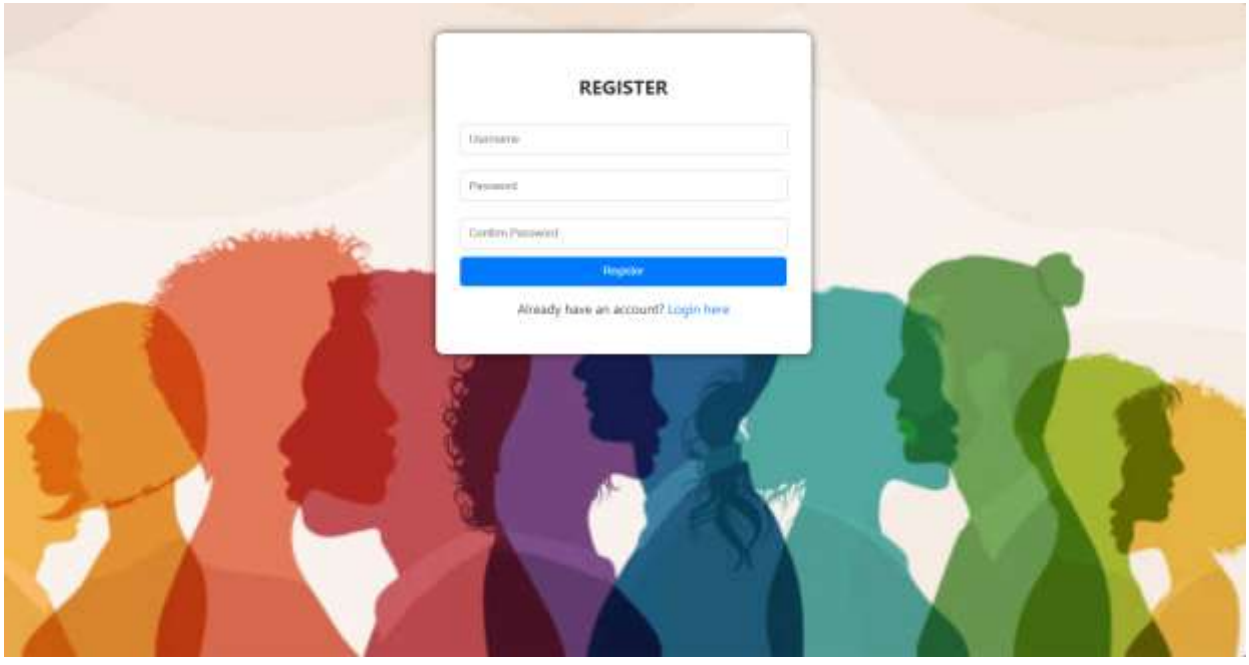


Fig 1. Register

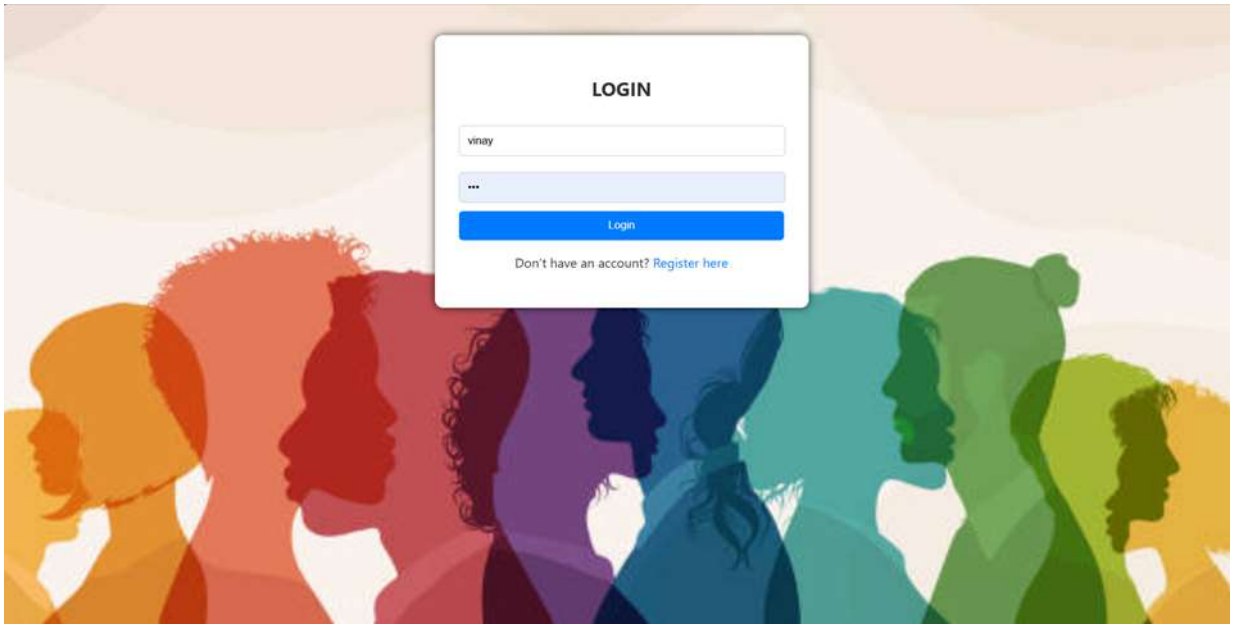


Fig 2. Login

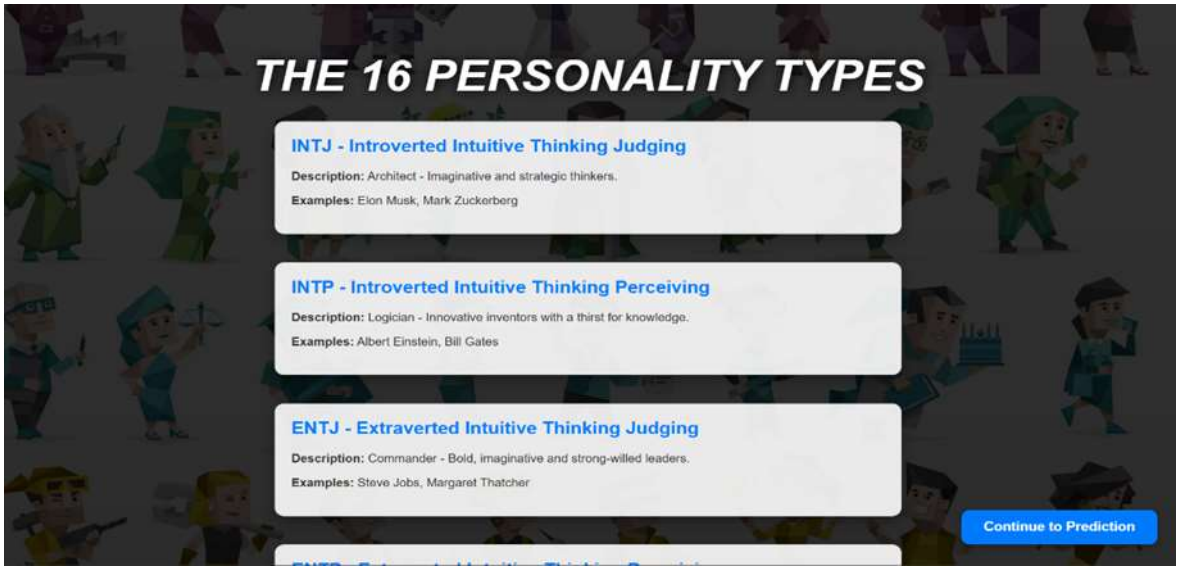


Fig 3. Basic Personality types

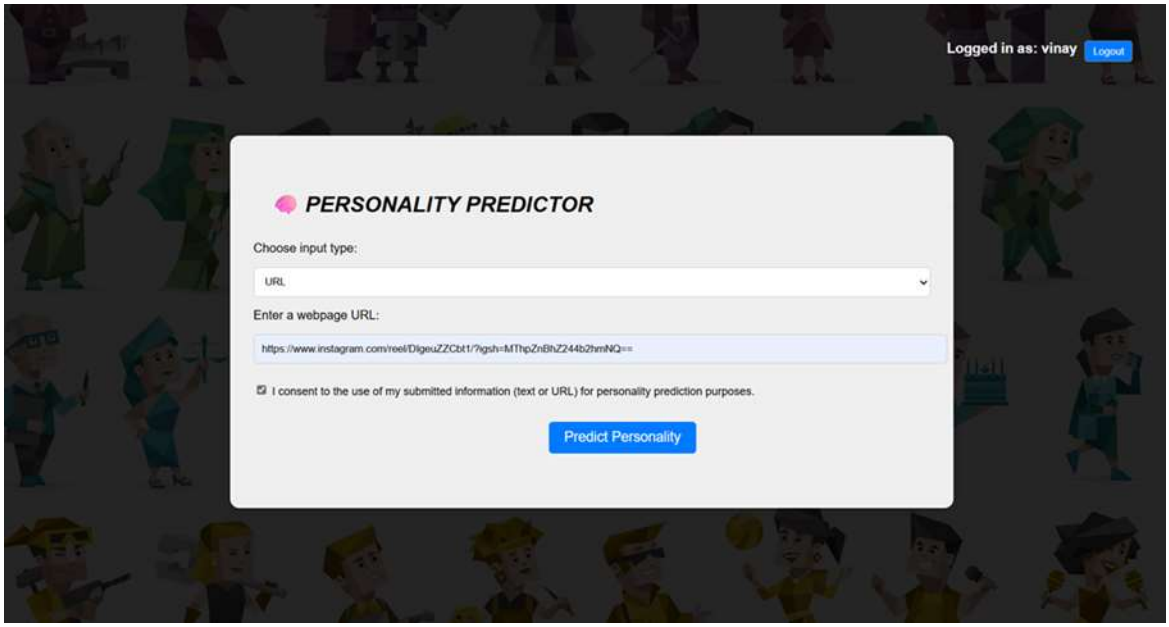


Fig 4. Personality Predictor

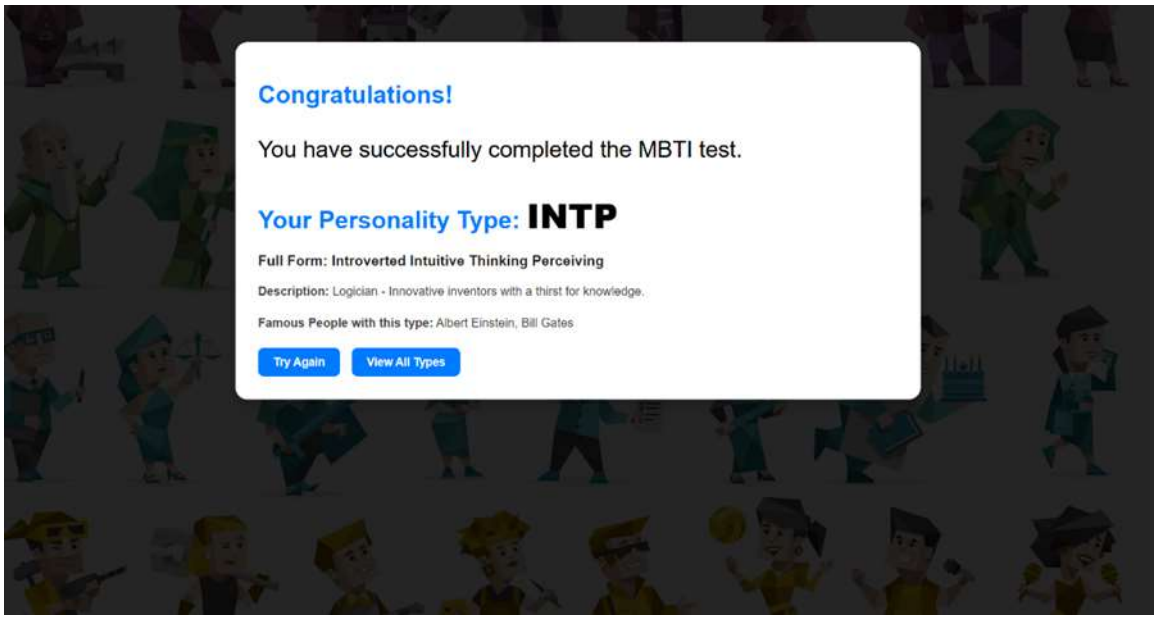


Fig 5. Predicted Message

7. CONCLUSION

The Personality Prediction System marks a significant step in leveraging machine learning and natural language processing techniques to derive psychological insights from user-generated text. The project successfully showcases the application of a pre-trained machine learning model based on the MBTI (Myers-Briggs Type Indicator) personality framework, providing a functional and interactive platform where users can input text and receive an accurate personality prediction.The system architecture was thoughtfully designed with a modular, layered approach, separating concerns across various components such as the frontend interface, backend processing, text preprocessing module, and machine learning layer.

This not only improved maintainability and scalability but also allowed for seamless data flow and easy integration of future enhancements. The use of Flask as a lightweight backend framework ensured quick handling of requests, while the machine learning model, trained and serialized using Python libraries, effectively predicted the user's personality type based on linguistic patterns in the provided text. During testing and deployment, the system demonstrated stable performance with acceptable accuracy, offering personality classifications that aligned with expected outcomes. The interactive web interface provided an intuitive and user-friendly experience, making it accessible even for non-technical users. The implementation of text preprocessing further enhanced model accuracy by cleaning and standardizing user input before classification.

8.FUTURE SCOPE

Although the current system performs well for basic personality prediction, there are several opportunities to enhance and expand its capabilities in future iterations:

Multi-language Support: Extend the system to process text inputs in multiple languages, broadening accessibility for global users.

Real-time Chat Analysis: Integrate the model into chat applications or social media platforms to analyze user personality in real time.

Emotion and Sentiment Detection: Combine personality prediction with emotion and sentiment analysis to offer deeper psychological insights.

Advanced Model Integration: Upgrade to more sophisticated NLP models such as BERT or GPT for improved prediction accuracy and contextual understanding.

User Authentication and Profile Management: Implement user login, result history, and profile dashboards for personalized user experiences.

Mobile Application Version: Develop a mobile-friendly version or dedicated app for on-the-go personality analysis. These future enhancements would further improve the system's practicality, accuracy, and user engagement potential.

9. REFERENCES

1. J. Golbeck, C. Robles, and K. Turner, "Predicting Personality with Social Media," *Proceedings of the 2011 Annual ACM Web Science Conference*, 2011, pp. 253–262. doi:10.1145/2527031.2527057.
2. Y. Bai, X. Zhou, J. Zhang, and J. Luo, "Predicting Personality Traits from Text Using Attention-Based Neural Networks," *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 139–148. doi:10.1145/3357384.3357890.
3. J. Pennebaker, M. Mehl, and K. Niederhoffer, "Psychological Aspects of Natural Language Use: Our Words, Our Selves," *Annual Review of Psychology*, vol. 54, no. 1, pp. 547–577, 2003. doi:10.1146/annurev.psych.54.101601.145041.
4. H. Mairesse, M. Walker, M. Mehl, and R. Moore, "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text," *Journal of Artificial Intelligence Research*, vol. 30, pp. 457–500, 2007.
5. F. L. Schmidt, "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings," *Psychological Bulletin*, vol. 124, no. 2, pp. 262–274, 1998.
6. A. Tadesse, H. Lin, B. Xu, and L. Yang, "Personality Predictions Based on User Behavior on the Facebook Social Media Platform," *IEEE Access*, vol. 6, pp. 61959–61969, 2018. doi:10.1109/ACCESS.2018.2870592.
7. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *arXiv preprint arXiv:1301.3781*, 2013.