

PREDICTING PERSONALITY TRAITS

Mini Project Report – Submitted by

VAIBHAVI SHENOY B

(4NM21CS198)

TVISHA PRAKASH

(4NM21CS195)

Under the guidance of

Ms. VAISHALI BANGERA

(Assistant Professor Gd-I)

In partial fulfillment of the requirements for the award of

the Degree of

Bachelor of Engineering in Computer Science Engineering

from

Visvesvaraya Technological University, Belagavi



N.M.A.M. INSTITUTE OF TECHNOLOGY

(An Autonomous Institution affiliated to Visvesvaraya Technological University, Belagavi)

Nitte – 574 110, Karnataka, India

APRIL 2024

CERTIFICATE

*This is to certify that the “Mini Project” submitted by **VAIBHAVI SHENOY B** bearing USN **4NM21CS198** of 6th semester B.E., a bonafide student of NMAM Institute of Technology, Nitte, has completed Mini Project on **PREDICTING PERSONALITY TRAITS** during **2023-2024** fulfilling the partial requirements for the award of degree of Bachelor of Engineering in **Computer Science & Engineering** at NMAM Institute of Technology, Nitte.*

Name and Signature of Mentor

Signature of HOD

ACKNOWLEDGEMENT

We dedicate this page to acknowledging and thanking those responsible for shaping the project. Without their guidance and help, the experience while constructing the dissertation would not have been so smooth and efficient.

First of all, I would like to thank our honorable principal **Dr. Niranjan N Chiplunkar** for their support and encouragement.

We owe our profound gratitude to **Dr Jyothi Shetty**, the Head of the Department of Computer Science and Engineering for providing this opportunity. Their dedication to academic excellence has greatly enriched my understanding and approach to this project. I am indebted to my esteemed mini project guide, **Ms. Vaishali Bangera**, for their expertise, patience, and commitment to my intellectual development. Their constructive feedback and guidance have significantly contributed to the refinement and success of this endeavor.

Lastly, I would like to express my heartfelt gratitude to my esteemed project team member for their unwavering dedication, collaboration, and professionalism. Their collective efforts and synergy have been instrumental in the realization of our project objectives.

ABSTRACT

Understanding human personality is crucial for various aspects of life, influencing career choices, relationships, and even mental health outcomes. The Myers-Briggs Type Indicator (MBTI) is a widely used framework that categorizes individuals based on four dichotomous preferences: Extraversion/Introversion, Sensing/Intuition, Thinking/Feeling, and Judging/Perceiving.

This work explores the potential of machine learning in predicting MBTI personality traits. We delve into the theoretical foundation of the MBTI framework, highlighting its four core dimensions and their influence on individual behaviour.

Furthermore, we explore the application of supervised learning algorithms in developing models capable of predicting MBTI types based on relevant data points. This research investigates the feasibility and effectiveness of utilizing machine learning to gain insights into individuals' personality through the lens of the MBTI framework.

By leveraging the power of machine learning, we aim to contribute to a deeper understanding of personality and its potential applications in various fields.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

CHAPTER 2 PROBLEM STATEMENT

CHAPTER 3 OBJECTIVES

CHAPTER 4 METHODOLOGY

CHAPTER 5 RESULTS

CHAPTER 6 FUTURE SCOPE

CHAPTER 7 CONCLUSION

CHAPTER 8 REFERENCES

INTRODUCTION

Understanding human personality is a fundamental pursuit in psychology, offering insights into the unique ways individuals think, feel, and behave. Personality traits act as the building blocks, shaping our characteristic patterns of thoughts, emotions, and actions. These traits are not merely fleeting moods but relatively stable tendencies that guide our responses to various situations and interactions. Investigating personality traits becomes crucial because they play a significant role in numerous aspects of our lives, influencing our choices, relationships, and even our success in different domains. By delving into the world of personality traits, we gain a deeper understanding of what makes us who we are and how these traits shape the tapestry of our individual experiences.

This project delves further by exploring the potential of machine learning models in predicting personality traits. By analyzing data points and patterns associated with these traits, we aim to develop a model capable of providing insights into an individual's personality based on specific inputs. This opens doors to various applications, potentially revolutionizing areas like personalized education, career guidance, and even tailored marketing strategies.

PROBLEM STATEMENT

Understanding human personality is crucial, as it shapes our thoughts, feelings, and behaviors. Personality traits act as the building blocks of these patterns, influencing our responses to various situations. This project aims to leverage the power of machine learning to predict individual personality traits. By analyzing data points and patterns associated with these traits, we aim to develop a model capable of providing insights into an individual's personality based on specific inputs. This opens doors to potential applications in personalized education, career guidance, and even tailored marketing strategies.

OBJECTIVES

This project aims to leverage the power of machine learning to develop a model capable of predicting individual personality traits. Utilizing supervised learning algorithms, we will train the model on a dataset consisting of labeled personality assessments and relevant user data. By evaluating the model's performance on unseen data, we intend to assess its accuracy in predicting personality characteristics, paving the way for potential applications in various fields.

The key aspects of this project are:

- **Focus on Machine Learning:** Clearly states the use of machine learning as the core technology for personality prediction.
- **Supervised Learning Approach:** Specifies the type of learning algorithm used to train the model.
- **Data-Driven Training:** Highlights the importance of a dataset with accurate personality assessments and relevant user data.
- **Evaluation and Accuracy:** Emphasizes the crucial step of evaluating the model's performance on unseen data to gauge its effectiveness.
- **Potential Applications:** Briefly mentions the broader implications of developing an accurate personality prediction model.

METHODOLOGY

1. Data Import

- Libraries: Necessary libraries like pandas and matplotlib are imported for data manipulation and visualization.
- Data Loading: The MBTI dataset is loaded from a CSV file named "mbti_1.csv". This file contains columns for MBTI types ("type") and text data associated with individuals ("posts").

2. Data Preprocessing

- Cleaning Text: A function named cleanText is defined to remove unnecessary characters like URLs, punctuation, and special symbols from the "posts" column. This helps standardize the text data for further analysis.
- Applying Cleaning Function: The apply method applies the cleanText function to each element in the "posts" column, ensuring consistent cleaning across all text data.
- Analyzing Personality Distribution: The frequency of each personality type is calculated using value_counts. This provides insights into the prevalence of different MBTI types within the dataset.
- Visualization: A bar chart is created using plt.bar to visualize the personality distribution, making it easier to understand the distribution of different types.

3. Feature engineering

Focuses on extracting potentially relevant information from the text data ("posts") related to different personality traits.

- Number of http links per comment: This feature captures the frequency of external links shared by individuals. It might be associated with openness to new information or a tendency to seek external validation.
- Music mentions per comment: This feature counts the occurrences of words related to music in the text. It could potentially indicate an interest in artistic expression or emotional engagement.

- Question marks per comment: The frequency of question marks might reflect a curious and inquisitive nature, seeking clarification or different perspectives.
- Image mentions per comment: This feature counts the occurrences of words related to images (e.g., "jpg"). It could be associated with a preference for visual communication or a more concrete thinking style.
- Exclamation marks per comment: The frequency of exclamation marks might indicate a more expressive or enthusiastic communication style, potentially reflecting extraversion or openness to experience.
- Ellipses per comment: Counting the occurrences of ellipses ("...") could be indicative of a more introspective or reflective personality, suggesting pauses in thought or open-ended statements.

4. Machine Learning Models

Implementation of three machine learning models for personality prediction:

- Logistic Regression: This linear model is a popular choice for classification tasks. It estimates the probability of a data point belonging to a particular class based on its features.
- K-Nearest Neighbors (KNN): This non-parametric model classifies a data point based on the majority class of its k nearest neighbors in the training data.
- Random Forest: This ensemble model combines multiple decision trees, each trained on a random subset of features and data points. This helps to reduce overfitting and improve generalization performance.
 - i. Data Splitting: The dataset is split into training and testing sets using `train_test_split`. This ensures that the models are evaluated on unseen data during testing.
 - ii. Model Building and Evaluation: Each model (Logistic Regression, KNN, Random Forest) is trained on the training data. The accuracy of each model is evaluated on the testing data using the score method. This provides an estimate of how well the model generalizes to unseen data.
 - iii. Individual Personality Trait Prediction: Building separate models for predicting each of the four personality dimensions (I/E, N/S,

T/F, J/P). This involves: Dropping the "type" and "posts" columns, keeping only the extracted features. Creating separate target variables (Y1, Y2, Y3, Y4) for each personality dimension. Following the same data splitting, model building, and evaluation steps as for overall personality prediction.

5. Visualizing Personality Traits: Unveiling Linguistic Patterns through MBTI-Specific Word Clouds

- **Leveraging Image Masks and Custom Words:** This section details the core methodology of using image masks specific to each MBTI personality type. These masks shape the word cloud, creating visually distinct representations. It explains the inclusion of custom word lists associated with each personality type, highlighting words particularly relevant to their characteristics.
- **Extracting Textual Features:** This delves into the process of extracting text data from the "posts" column within your dataset. It explains the specific features chosen to analyze, such as the frequency of links, music mentions, punctuation marks, etc., and how they might relate to different personality traits.
- **Word Cloud Generation and Visualization:** This section focuses on the technical aspects of generating word clouds using the WordCloud library. It details parameters like background color, maximum word count, stopword removal, and the chosen color function. It also explains the visualization process of displaying the generated word clouds alongside their corresponding image masks for comparison.
- **Insights and Interpretation:** This discusses potential interpretations of the generated word clouds. It explores how the prominent words within each cloud reflect the characteristics of the associated MBTI type.

6. Predicting MBTI Traits using Machine Learning

This section dives into the detailed steps of a machine learning system for predicting MBTI personality traits based on text data

- **User Input:** The code incorporates a loop that allows users to input sentences or text. The user input is processed through the same TF-IDF vectorizer used for the training data, converting it into a numerical representation.

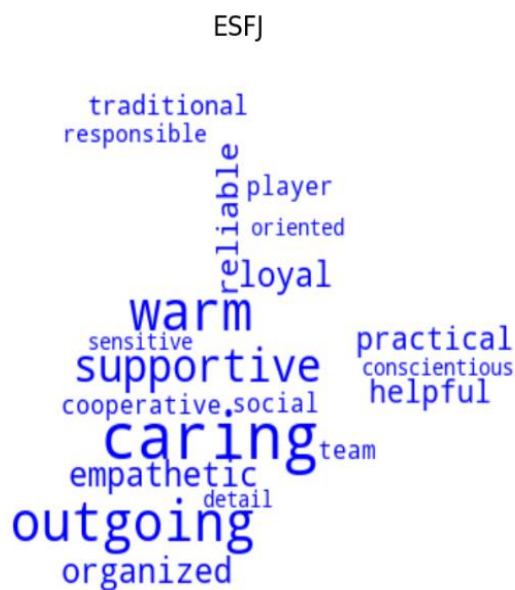
- Prediction: The trained model predicts the MBTI personality trait for the user-provided text based on the learned patterns from the training data.

RESULTS

Personality Trait	Logistic Regression	KNN	Random Forest
Introversion/Extraversion	76.94%	83.10%	99.29%
Sensing/Intuition	86.22%	88.60%	99.39%
Thinking/Feeling	62.93%	78.36%	98.60%
Judging/Perceiving	60.31%	77.12%	98.58%

- Random Forest consistently achieved the highest accuracy across all personality traits, indicating its strong performance in predicting these categories
- KNN generally performed better than Logistic Regression, except for the Thinking/Feeling trait.
- The accuracy for Introversion/Extraversion and Sensing/Intuition is significantly higher than for Thinking/Feeling and Judging/Perceiving.

Among the models tested, Random Forest consistently achieved the highest accuracy, suggesting its strong performance in predicting these categories. This makes it a suitable choice for predicting personality traits based on the provided data.



Text data analysis provides valuable insights into the characteristics associated with the ESFJ personality type. The identified words support the established understanding of ESFJs as social, empathetic, and organized individuals who prioritize tradition and social connection. Similarly , text data analysis is done for all other types of personality traits.

```
Enter a sentence (or type 'quit' to exit): I enjoy spending time alone exploring new ideas
```

```
Predicted personality trait: INTP
```

```
Enter a sentence (or type 'quit' to exit): I often find myself lost in thought, contemplating philosophical questions about the meaning of life and the universe
```

```
Predicted personality trait: INFP
```

```
Enter a sentence (or type 'quit' to exit): quit
```

Overall, this approach demonstrates the development of a personality prediction model using a Random Forest classifier and TF-IDF feature extraction. The code provides a framework for training the model on a dataset of text and personality labels, and then using it to predict personality traits for new user inputs.

FUTURE SCOPE

Predicting personality traits holds immense potential for the future in various ways-

- **Understanding Human Behavior:** Research within this field can contribute to a deeper understanding of how personality influences our thoughts, behaviors, and choices. This knowledge can be valuable for various fields, including psychology, sociology, and marketing.
- **Teachers Understanding Students:** By understanding a student's personality traits, teachers can tailor their teaching methods to cater to individual needs and learning styles.
- **Optimal Group Formation:** Personality prediction can be used to create well-balanced teams within the classroom leading to more effective collaboration and project outcomes.
- **Career Guidance and Job Matching:** Personality assessments could help individuals identify suitable career paths and connect them with jobs that align with their strengths and preferences. This could lead to greater job satisfaction and productivity.
- **Personalized Experiences:** Social media platforms could suggest content you'd find engaging, e-commerce platforms could recommend products best suited to your preferences, and educational platforms could adapt their teaching methods to your learning style.

Overall, the future of personality prediction holds great promise for personalized experiences, enhanced human-computer interaction, and deeper insights into human behavior. As the technology evolves with careful consideration of ethical concerns, it can become a powerful tool for self-understanding, career guidance, and overall well-being.

CONCLUSION

Predicting personality traits holds immense potential to reshape the way we interact with ourselves and others, offering a deeper understanding of human behavior and paving the way for a more personalized and supportive future. By analyzing data points like text, social media behavior, and even smartphone usage, we can gain valuable insights into someone's openness to new experiences, level of conscientiousness, and tendencies toward extroversion or introversion. This knowledge can be harnessed in various ways:

Education: Teachers can tailor their teaching methods to individual learning styles, identify students who might benefit from additional support, and create well-balanced teams for collaborative projects. This can lead to a more effective and engaging learning environment for all students.

Career Guidance: Individuals can gain valuable insights into potential career paths that align with their natural strengths and preferences, leading to greater job satisfaction and productivity.

Mental Health: Early identification of personality traits linked to potential mental health concerns can enable proactive interventions and support services.

Human-Computer Interaction: AI assistants and virtual companions can become more attuned to our personalities, leading to more natural and personalized interactions.

In conclusion, the ability to predict personality traits offers a powerful tool for self-understanding, personalized experiences, and deeper insights into human behavior. As this technology continues to evolve with careful consideration of ethical implications, it has the potential to revolutionize various aspects of our lives, creating a future where we can leverage our unique personalities to thrive in education, careers, and personal well-being.

REFERENCES

- i. <https://ieeexplore.ieee.org/document/10199294>

