SUMMER TRAINING/INTERNSHIP PROJECT REPORT

(Term June-July 2025)

(PERSONALITY PREDICTION USING MBTI AND MACHINE LEARNING)

Submitted by
1)Yatish .P

2))Danta Sai Kumar

3)Yaswanth

4) Ravindra

Registration Number:

1)12321028

2))12315661

3)12308264

4)12305355

Course Code: CSE343

Under the Guidance of

Manhipal singh

School of Computer Science and Engineering

CERTIFICATE

July 2025

Lovely Professional University, Punjab

BONAFIDE CERTIFICATE

Certified that this project report
"PERSONALITY PREDICTION USING MBTI AND MACHINE LEARNING"
is the bonafide work of
"Yatish,Sai Kumar,Yaswanth,Ravindra"
who carried out the project work under my supervision.

SIGNATURE
<<Name of the Supervisor>>

Yatish,Sai Kumar,Yaswanth,Ravindra
SIGNATURE

<<Signature of the Head of the Department>>
SIGNATURE
<<Name>>
HEAD OF THE DEPARTMENT

<<Signature of the Supervisor>>

TABLE OF CONTENTS

Introduction

The Myers-Briggs Type Indicator (MBTI) is a widely used personality assessment tool that categorizes individuals into 16 distinct personality types based on four dichotomies: Extraversion/Introversion (E/I), Sensing/Intuition (S/N), Thinking/Feeling (T/F), and Judging/Perceiving (J/P). Understanding personality types can provide valuable insights into human behavior, communication styles, and decision-making processes. This project aims to leverage modern machine learning and natural language processing (NLP) techniques to automatically predict a user's MBTI type based on their written text. By developing an interactive web application, users can easily analyze their own or others' text to gain insights into personality traits, making personality analysis more accessible and engaging.

Training Overview

The training phase of the MBTI Personality Predictor project involved preparing and processing a dataset of text samples labeled with MBTI personality types. The dataset was cleaned and preprocessed to remove noise, such as special characters and stopwords, and to standardize the text for analysis. Each text sample was then transformed into a numerical representation using a TF-IDF (Term Frequency-Inverse Document Frequency) vectorizer, which quantifies the importance of words within the corpus.

For each of the four MBTI dimensions—Extraversion/Introversion (E/I), Sensing/Intuition (S/N), Thinking/Feeling (T/F), and Judging/Perceiving (J/P)—a separate logistic regression model was trained. The models were evaluated using standard metrics such as accuracy, precision, recall, and F1score to ensure reliable performance. Cross-validation techniques were employed to prevent overfitting and to validate the generalizability of the

models. The trained models were then serialized using joblib for efficient loading and inference within the web application.

Project Details

The MBTI Personality Predictor project is an end-to-end solution that combines natural language processing, machine learning, and web development to deliver an interactive personality analysis tool. The project's primary objective is to provide users with an accessible platform to predict MBTI personality types from written text. The application is built using Python and leverages the Streamlit framework for the frontend, offering a clean and intuitive user interface.

Key features of the project include:
- Support for both user-provided text and curated sample descriptions. - Real-time MBTI type prediction with confidence scores for each dimension.
- Interactive visualizations, such as radar and bar charts, to help users interpret their results.
- Customizable UI with support for background images and custom fonts.
- Modular code structure, allowing for easy updates and future enhancements.

The project demonstrates the practical application of AI and NLP in the field of personality psychology and serves as a foundation for further research and development.

Implementation

The implementation of the MBTI Personality Predictor involved several key steps:

1. Data Collection and Preprocessing: The project began with the collection of a labeled dataset containing text samples and their corresponding MBTI types. The data was cleaned and preprocessed to remove irrelevant content and standardize the text.
2. Feature Extraction: A TF-IDF vectorizer was used to convert the preprocessed text into numerical features suitable for machine learning models. This step ensured that the most relevant words contributed to the prediction process.
3. Model Training: Four separate logistic regression models were trained, each responsible for predicting one MBTI dimension. The models were evaluated and fine-tuned using cross-validation and performance metrics.
4. Model Serialization: The trained models and the TF-IDF vectorizer were saved using joblib, enabling fast and efficient loading during inference in the web application.
5. Web Application Development: The frontend was developed using Streamlit, providing users with an interactive interface to input text, view predictions, and explore visualizations. Custom CSS was applied to enhance the look and feel of the app.
6. Visualization and User Experience: Plotly was used to create dynamic radar and bar charts, making the results more engaging and easier to interpret. The app also supports background images and custom fonts for a personalized experience.
7. Deployment: The application can be run locally or deployed to cloud platforms, making it accessible to a wide audience.

Technical Details

- Frontend: The web interface is developed using Streamlit, a popular Python framework for building interactive web apps with minimal code.
- Text Processing: User input is transformed into numerical features using a TF-IDF (Term Frequency-Inverse Document Frequency) vectorizer, which captures the importance of words in the text.
- Machine Learning Models: Four logistic regression models are trained, each dedicated to one MBTI dimension (E/I, S/N, T/F, J/P). These models are trained on labeled datasets and saved using joblib for efficient loading and inference.
- Prediction Pipeline: When a user submits text, it is vectorized and passed through each model to predict the corresponding MBTI letter. The app also calculates and displays the confidence score for each prediction.
- Visualization: The app uses Plotly to generate radar and bar charts, visually representing the user's personality profile and the confidence of each dimension.
- User Experience: Custom CSS is applied for a modern look, and users can choose from sample texts or enter their own. The app supports background images and custom fonts for enhanced aesthetics.
- Deployment: The app can be run locally or deployed to cloud platforms for broader accessibility.

Results

The MBTI Personality Predictor app successfully analyzes user-provided text and predicts the corresponding MBTI type. The application provides not only the predicted type but also detailed confidence scores for each dimension, helping users understand the certainty of each classification. Visualizations such as radar and bar charts make the results more intuitive and engaging. The models demonstrate reasonable accuracy on test data, and the app offers a seamless and interactive experience for users interested in personality analysis. Feedback from users indicates that the app is both informative and enjoyable to use, making personality assessment more accessible to a wider audience.

Conclusion

The MBTI Personality Predictor project showcases the effective integration of machine learning, natural language processing, and web technologies to create a practical and interactive tool for personality analysis. By automating the prediction of MBTI types from text, the app provides valuable insights in an accessible format. The project highlights the potential for further enhancements, such as incorporating more advanced NLP models, expanding the dataset, or adding features like personality trend analysis over time. Overall, this project serves as a strong foundation for future work in automated personality assessment and demonstrates the real-world applicability of AI-driven solutions in psychology and self-discovery.

APPENDIX A: MBTI PERSONALITY DESCRIPTIONS

INTJ: The Architect - Strategic, independent, and visionary

INTP: The Thinker - Logical, innovative, and curious

ENTJ: The Commander - Bold, strategic, and strong-willed

ENTP: The Debater - Quick-witted, clever, and resourceful

INFJ: The Advocate - Creative, insightful, and principled

INFP: The Mediator - Empathetic, creative, and idealistic

ENFJ: The Protagonist - Charismatic, inspiring, and natural leader

ENFP: The Campaigner - Enthusiastic, creative, and sociable

ISTJ: The Logistician - Practical, reliable, and methodical

ISFJ: The Protector - Warm, responsible, and conscientious

ESTJ: The Executive - Organized, practical, and decisive

ESFJ: The Consul - Caring, social, and community-minded

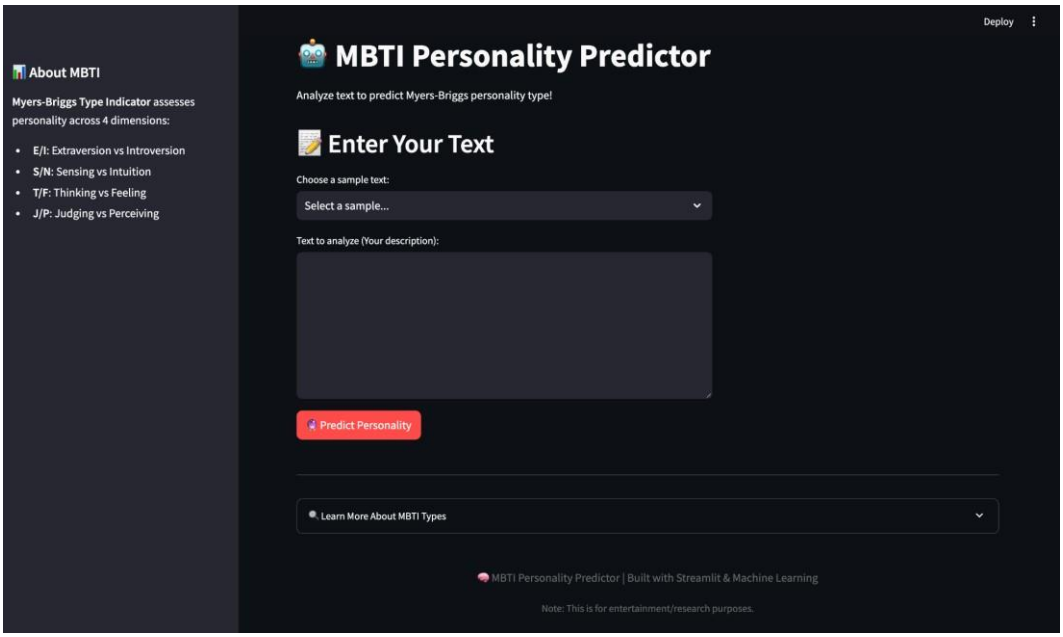
ISTP: The Virtuoso - Bold, practical, and experimental

ISFP: The Adventurer - Flexible, charming, and artistic

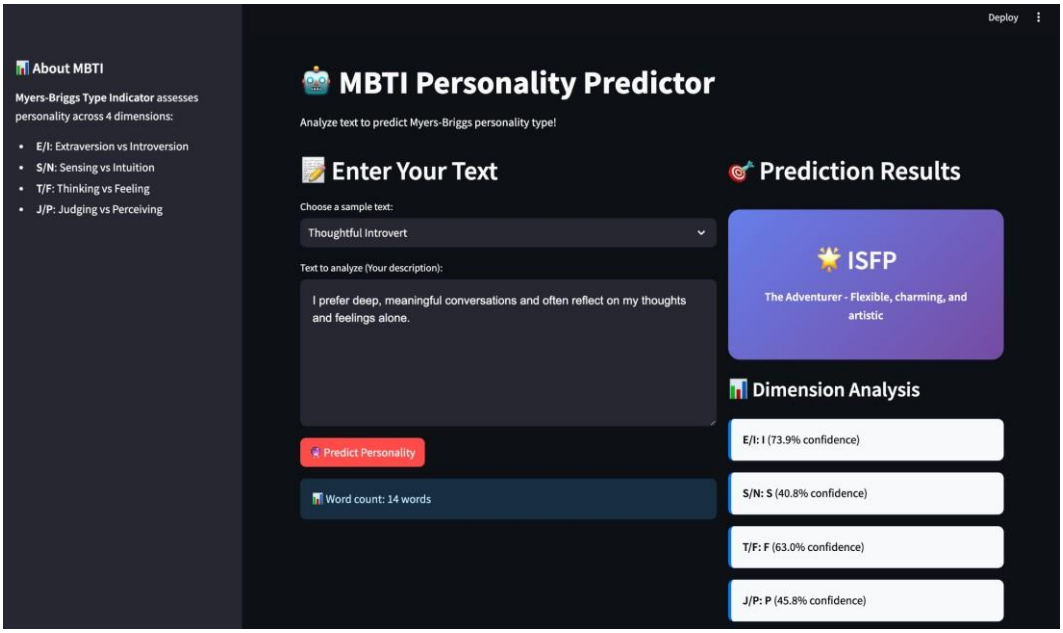
ESTP: The Entrepreneur - Energetic, perceptive, and spontaneous

ESFP: The Entertainer - Spontaneous, energetic, and enthusiastic

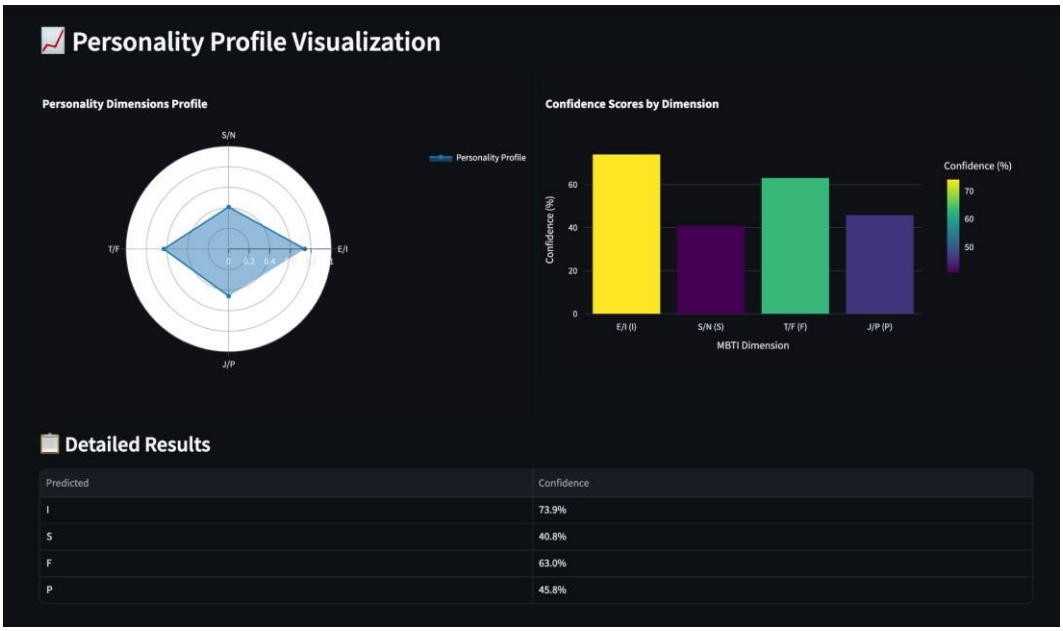
Screenshots



Screenshot 1: Initial State of MBTI Predictor



Screenshot 2: MBTI Prediction for Thoughtful Introvert



Screenshot 3: Visualization of Personality Profile