**Personality-Trait-Analysis**

*A Project Report*

*Submitted by*

**Raghav Kwatra – 2214110584**

**Anusha Anand - 2114110004**

**Gaurav Bajaj - 2114110006**

**Harshita Jain – 2114110062**

*In partial fulfilment of the requirement*

*For Project Based Learning of*

**BACHELOR IN TECHNOLOGY**

*In*

# COMPUTER SCIENCE AND ENGINEERING

*For*

*Sub: Deep Learning*

*Under the guidance of*

*Prof. Nisha Auti*



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING BHARATI VIDYAPEETH (DEEMED TO BE UNIVERSITY) COLLEGE OF ENGINEERING, PUNE-43 2024-25**

# BHARATI VIDYAPEETH (DEEMED TO BE UNIVERSITY) COLLEGE OF ENGINEERING, PUNE- 43



**CERTIFICATE**

This is to certify that the Project Based Learning report titled **Personality-Trait-Analysis**, submitted by **Raghav Kwatra (2214110584), Anusha Anand (2114110004), Gaurav Bajaj (2114110006), Harshita Jain (2114110061),** to the Bharati Vidyapeeth (Deemed to be University), College of Engineering, Pune - 43 for the award of the degree of **BACHELOR OF TECHNOLOGY** in Computer Science and

Engineering is a bonafide record of the PBL work done by them under my supervision.

Place: Pune Name of the subject teacher:

Prof. Nisha Auti

Date:

**Index**

|  |  |  |
| --- | --- | --- |
| **Sr no.** | **Title** | **Page no.** |
| 1. | Abstract | 4 |
| 2. | Introduction | 5 |
| 3. | Literature Study of the Topic | 6 |
| 4. | Problem Statement | 7 |
| 5. | Objectives | 8 |
| 6. | Methodology Used | 9 |
| 7. | System Architecture | 13 |
| 8. | Result and Analysis | 14 |
| 9. | Conclusion | 15 |

**Abstract**

This project aims to classify personality traits based on the Myers-Briggs Type Indicator (MBTI) using natural language processing (NLP) and deep learning techniques. By analyzing a user's social media posts, we seek to predict personality types through various machine learning models. The project leverages the MBTI dataset, containing users' text-based data labeled with their MBTI types, and consists of preprocessing and three model implementations: Naive Bayes, Support Vector Machine (SVM), and Convolutional Neural Network (CNN). Each model's performance is compared to identify the most accurate and effective approach. NLP preprocessing techniques, such as tokenization and TF-IDF, transform the data into features suitable for model training. Our results demonstrate that deep learning models, especially CNNs, significantly enhance the accuracy and robustness of personality prediction, highlighting the effectiveness of deep learning in understanding complex, human-centric data. This project could lead to advancements in personalized recommendations, mental health analysis, and other applications where understanding personality is valuable.

**Introduction**

The rapid growth of social media has created an unprecedented amount of text data, which can reveal users' personalities. Personality analysis based on the MBTI (Myers-Briggs Type Indicator) is particularly relevant for applications in psychology, marketing, and mental health. MBTI divides personality into 16 types based on four dichotomies: Extraversion-Introversion, Sensing-Intuition, Thinking-Feeling, and Judging-Perceiving. This project explores automated MBTI classification from social media text data using NLP and machine learning. We employ and compare three models: Naive Bayes, SVM, and CNN, each representing an increasing level of complexity in NLP processing. This analysis seeks to improve personality prediction accuracy and understand which model best captures the subtleties of personality traits expressed in text. By achieving this, the project has implications for personalized user experiences and potential uses in human-centered AI systems.

**Literature Study**

 Background **on Personality Analysis**:

* Psychological personality analysis, especially MBTI, has been widely used in various fields.

 Early **Computational Approaches**:

* Traditional machine learning models like Naive Bayes and SVM for text analysis.
* Reliant on feature engineering (e.g., Bag of Words, TF-IDF).

 NLP **and Machine Learning Advancements**:

* NLP developments enable better text-based personality analysis.
* Deep learning (CNNs, RNNs) can capture complex patterns without manual feature engineering.

 Emerging **Techniques**:

* Transformer models (e.g., BERT) improve semantic understanding, allowing more accurate personality insights.
* Deep learning, particularly CNN, has shown promise in MBTI classification but is still under-explored.

**Problem Statement**

The objectives of this project include:

1. To preprocess text data from the MBTI dataset for effective NLP application.
2. To implement and evaluate a Naive Bayes classifier for personality classification.
3. To apply and compare a Support Vector Machine (SVM) for its effectiveness in capturing linear separability in MBTI data.
4. To develop a CNN model capable of capturing more complex language patterns associated with personality types.
5. To analyze and compare the performance of these models in terms of accuracy, precision, and recall.

**Objectives**

The objectives of this project are:

* To develop a deep learning-based system that performs user-interest-based video summarization.
* To implement CNN-based object detection to accurately identify objects in video frames.
* To use transformer-based models in NLP to process user queries and map them to detected objects.
* To generate video summaries that are contextually relevant to the user's query.
* To optimize the deep learning models for real-time summarization and user interaction.

**Methodology**

 **Data Preprocessing**:

* Import MBTI dataset, clean and standardize text.
* Tokenize and apply TF-IDF for converting text to numeric features.
* Code :

 **Naive Bayes Model**:

* Apply Multinomial Naive Bayes to preprocessed features.
* Train and evaluate on test data.
* Code

 **SVM Model**:

* Use linear SVM for class separation in feature space.
* Optimize hyperparameters for best results.
* Code

 **CNN Model**:

* Convert text to embeddings (e.g., Word2Vec, GloVe).
* Apply convolutional layers to capture language patterns.
* Pooling layers for dimensionality reduction and feature extraction.
* Code:

 **Evaluation Metrics**:

* Compare models using accuracy, precision, recall, and F1-score.

**System Architecture**

The architecture is organized as a sequential pipeline where each model operates on the same preprocessed dataset. Data preprocessing is applied consistently to ensure uniform input across the Naive Bayes, SVM, and CNN models. Each model has a separate training and testing phase, with data passed through distinct architectures:

1. **Naive Bayes Model**: Inputs preprocessed text vectors and outputs personality type probabilities.
2. **SVM Model**: Similar input with a focus on hyperplane-based classification for MBTI types.
3. **CNN Model**: Utilizes word embeddings as input and applies convolutional layers to identify deeper text patterns.
4. **Data Preprocessing**:

* Standardizes dataset input across all models.

1. **Model Architecture**:

* **Naive Bayes**: Applies probability-based text classification.
* **SVM**: Creates a hyperplane for text class separation.
* **CNN**: Uses embeddings with convolutional layers for advanced pattern recognition.

1. **Output**:

* Models are trained and tested to produce performance metrics.
* Results highlight model comparison and performance.

**Results and Analysis**

Upon evaluating each model, it was observed that the Naive Bayes classifier provided a baseline accuracy but struggled with complex language patterns, particularly in handling class imbalances and subtleties in text indicative of personality traits. The SVM model showed an improvement, achieving better precision and recall across most MBTI classes. This can be attributed to SVM's robustness in handling high-dimensional data and its ability to find a hyperplane that maximally separates classes.

The CNN model outperformed both the Naive Bayes and SVM models, achieving the highest accuracy, likely due to its ability to capture word semantics via embeddings and detect non-linear relationships through its convolutional layers. The CNN model demonstrated a particular advantage in recognizing personality nuances, capturing implicit language cues, and significantly reducing misclassifications. These results suggest that deep learning models, particularly CNNs, offer a more nuanced approach to personality classification in text-based datasets.

**Conclusion**

This project illustrates the effectiveness of deep learning models in personality classification based on text data. While traditional methods like Naive Bayes and SVM provide reasonable accuracy, CNNs excel in capturing the complexities of human language and improving personality prediction accuracy. The successful implementation of this system can pave the way for applications in digital psychology, personalized marketing, and human-centered AI systems. Future work could incorporate additional data sources and advanced NLP techniques, such as transformers, to further enhance model accuracy and generalizability in real-world applications.