VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2024, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 12, Number 3, July-September 2024 pp:35-43

Enhancing MBTI Personality Prediction from Text Data with Advance Word Embedding Techniques

Nimra Ashraf [1], **Rao Sohail Iqbal** [2], **Shehar Bano** [1], **Hafiz Muhammad Azeem** [3*], **Shagufta Naz** [1]

[1]M.S Scholar, Department of Computer Science. Government College University, Faisalabad, Pakistan; [2]Lecturer, Department of Computer Science. Government College University, Faisalabad, Pakistan; [3]Lyallpur Business School, Government College University, Faisalabad, Pakistan

Keywords: *Personality Prediction, Myers-Briggs Type Indicator (MBTI), Machine Learning Classifiers, Text Data Analysis, GloVe and BERT Embeddings.*

Journal Info:
Submitted:
Aug 01, 2024
Accepted:
Aug 16, 2024
Published:
Aug 20, 2024

Abstract

Understanding human personality traits is crucial for various domains, including psychology, education, and human resources. The Myers-Briggs Type Indicator (MBTI) is a widely recognized psychological assessment tool, categorizing individuals into one of sixteen distinct personality types. The existing methodologies, which primarily relied on Word2Vec embeddings and traditional machine learning models. This research focused on enhancing Myers-Briggs Type Indicator (MBTI) personality prediction from text data through advanced word-embedding techniques, specifically GloVe and BERT. The research investigates the effectiveness of various Machine Learning Classifiers, including Random Forest, XGBoost, LinearSVC, SGD, Logistic Regression, and CatBoost, in predicting MBTI personality types. Additionally, the impact of preprocessing techniques such as text cleaning, tokenization, TF-IDF vectorization, GloVe, and BERT embeddings on classification performance is examined. Furthermore, the research explores strategies for addressing class imbalance through upsampling techniques. Results indicate high accuracy and performance across multiple classifiers, with XGBoost achieving the highest accuracy of 97.33%.

***Correspondence author email address:** hmazeem@gcuf.edu.pk
DOI: 10.21015/vtse.v12i3.1864

1 Introduction

Personality encompasses the unique ways individuals think, feel, and behave, distinguishing one person from another. With the evolution of society, diverse methodologies have emerged to understand individuals based on their social interactions and linguistic expressions. Social media platforms such as Facebook, Instagram, Twitter, YouTube, TikTok, and Weibo have become prevalent among internet users, offering rich data sources for studying and discerning individuals' online behaviors, preferences, and personalities. Users engage with various aspects of their lives on social media, spanning family connections, psychological well-being, financial matters, societal relations, and political discourse, thereby providing researchers with a wealth of data for analysis.

Social media platforms have become instrumental in predicting personality traits due to their abundance of user-generated content. Users' online behaviors, language use, and posted content offer valuable insights into their personalities. With billions of active users worldwide, social media platforms have become significant arenas for individuals to express themselves and engage with diverse topics, thereby providing researchers with ample opportunities to discern patterns and anticipate personality traits [1].

Computational personality recognition has gained traction within the realms of affective computing and artificial intelligence [2]. Notably, the Myers-Briggs Type Indicator (MBTI) and the Big Five model have emerged as prominent models for personality assessment. While the Big Five model has garnered greater acceptance among researchers, the MBTI has faced scrutiny in scientific circles [3]. This research focus on utilizing an MBTI dataset to delve into the intricate realm of personality traits, aiming to uncover insights into how individuals' online behaviors and language usage correlate with their MBTI personality types.

The Myers-Briggs Personality system, conceptualized by Katherine Cook Briggs and Isabel Briggs Myers, draws its foundations from Carl Jung's psychological theories 1923 [4]. The MBTI provides a structured framework for categorizing individuals into distinct personality types based on four key classification categories: Introversion or Extroversion, Judging or Perceiving, Thinking or Feeling, and Sensing or Intuition. These categories combine to yield 16 unique personality types, each characterized by a specific combination of traits. Leveraging natural language processing and machine learning techniques to predict MBTI personality types from text data has emerged as a promising avenue for understanding and interpreting personality traits.

The MBTI holds significance across multiple disciplines, including psychology, education, organizational development, career counseling, and personal development. In psychology, the MBTI aids in understanding human behavior and thinking patterns, guiding individuals towards self-awareness and improved communication [5]. Within educational contexts, the MBTI informs pedagogical practices and fosters personalized learning environments [6]. In organizational settings, the MBTI facilitates team dynamics, communication enhancement, and employee performance optimization [7]. Career counselors utilize the MBTI to guide individuals towards suitable career paths aligned with their personality preferences and interests [8]. On a personal level, the MBTI serves as a tool for self-awareness, self-reflection, and personal growth, aiding individuals in understanding their communication styles and social interactions [9].

[10] enhanced data quality and reduced SMOTE-generated noise resulting from inaccuracies. Preprocessing actions included data cleaning, lemmatization, tokenization and word embedding with Word2Vec. The CBOW model was chosen over Skip-gram. Data was split for training and testing, and six machine learning models were employed: Logistic Regression, Linear Support Vector Classification, Stochastic Gradient Descent, Random Forest, Extreme Gradient Boosting, and CatBoost. Data imbalance was addressed using SMOTE.

This research focused on enhancing Myers-Briggs Type Indicator (MBTI) personality prediction from text data through advanced word-embedding techniques, specifically GloVe and BERT. The data preprocessing phase involves critical tasks like text cleaning, tokenization, and converting MBTI-type labels into binary vectors representing the four dichotomous dimensions: I/E, N/S, F/T, and J/P. Exploratory Data Analysis (EDA) examines the distribution of posts across MBTI types, analyzing sentence length distribution and token frequency to provide insights into the dataset's structural composition and prevalent themes. Multiple approaches, including TF-IDF vectorizer, GloVe and BERT word embedding techniques, along with various machine learning classifiers, such as Random Forest, XGBoost, LinearSVC, SGD, Logistic Regression and CatBoost, are explored to improve MBTI personality prediction. The subsequent sections outline the research framework, detailing data preprocessing techniques, word embedding implementation, and classifier performance evaluation, aiming to advance MBTI personality prediction using state-of-the-art machine learning techniques.

2 Related Word

User personality recognition poses challenges due to small sample sizes and imbalanced data distribution. [11] proposed PSO-SMOTETomek, a method combining particle swarm optimization (PSO) with synthetic minority oversampling techniqueTomek Link (SMOTE-Tomek) resampling, enhancing classification accuracy by up to 10 percent. [12] explored predicting MBTI personality types based on scraped labeled texts using pre-trained language models. Their study achieved promising accuracy and investigated the use of fine-tuned BERT models for personality-specific language generation.

[13] employed development tools for automating personality type prediction, achieving promising results with the Gradient Boosting model implemented with XGBoost. [14] employed the k-means clustering algorithm to classify personality types, achieving effective unsupervised learning. [15] explored the relationship between personality types and dream experiences, revealing differences in dream characteristics based on MBTI dimensions.

[16] predicted personality traits from social media posts, achieving accurate predictions for MBTI dimensions in the Turkish language. [17] predicted personality scores on the MBTI scale using various machine learning models and feature extraction methods. [18] reviewed the utilization of MBTI in medical education, highlighting its potential benefits in improving communication skills and addressing biases. [19] proposed a data-centric approach utilizing NLP to predict personality types based on MBTI, demonstrating outstanding performance with the XGBoost classifier. [20] classified human personalities based on the MBTI method using tweets, achieving high accuracies with Support Vector Machine (SVM) classifiers, and accuracies up to 84.07% were achieved. [21] utilized a Random Forest classifier to predict personality types based on MBTI, outperforming other machine learning algorithms.

[8] presented an advanced system for personality classification and professional profile prediction, leveraging data mining techniques. [22] employed machine learning techniques to predict MBTI types, achieving notable accuracy rates and demonstrating the feasibility of predicting personality types from social media content. [23] proposed a data-centric approach for predicting personality types based on MBTI, emphasizing the importance of data quality and representativeness. [24] investigated personality detection in artificial intelligence using Transformer models, outperforming traditional methods. [25] conducted a comprehensive analysis of classical machine learning algorithms within the MBTI framework, showcasing efficient outcomes for personality trait prediction.

[26] employed machine learning methods to predict personalities based on textual data, integrating the CRISP-DM framework with agile methodology for efficient development cycles. [27] investigated mouse operational behaviors as a novel data source for online personality assessment, achieving promising recognition accuracies. [28] explored gender differences in personality traits among software engineers, highlighting implications for diversity and inclusion initiatives. [29] focused on predicting Myers–Briggs Type Indicator (MBTI) personality types of Twitter users using Long Short-Term Memory Networks (LSTM) and explored personality traits alongside career suitability. [3] compared linguistic similarities between individuals labeled with MBTI types and Big Five dimensions, offering insights into MBTI's influence on language use. [30] developed a web application using Flask and machine learning models to predict personality traits from social media text data, demonstrating the effectiveness of oversampling techniques. [31] investigated the use of pre-trained language models like BERT for MBTI classification, showcasing improved performance compared to traditional methods. [32]

combined word-embedding techniques with ensemble methods to predict MBTI personality types, offering a language-agnostic approach adaptable to different languages.

[9] employed NLP and machine learning classifiers to predict personality types and suggest career paths based on MBTI types. [33] discussed the influence of personality traits on problem-solving approaches and presented a method for predicting personality traits from smartphone typing data. [34] proposed a CA personality model for AI-based conversational agents, examining its influence on social and functional tasks. [35] classified users into profiles for customizing game-based learning environments, providing insights for game designers and researchers[35].

3 Methodology

The Myers-Briggs Type Indicator (MBTI) categorizes individuals into 16 personality types based on four dimensions: Introversion-Extroversion, Intuition-Sensing, Thinking-Feeling, and Judging-Perceiving. Widely used in business, research, and personal development, it offers insights into preferences and behaviors. Despite criticisms about its reliability, the MBTI remains influential. This research explores the link between MBTI types and writing styles to assess its effectiveness in behavior analysis.

3.1 Data Preprocessing

Data preprocessing is a vital step in preparing textual data for machine learning tasks, ensuring the dataset is clean, consistent, and suitable for model training. This involves several stages, starting with text cleaning and tokenization, where unwanted elements like URLs, emoticons, and non-alphanumeric characters are removed. Text is then normalized, and word tokenization is performed to structure the data appropriately for natural language processing tasks. Filtering is applied to remove irrelevant words, while ensuring that important delimiters are preserved to maintain data integrity.

Subsequently, advanced preprocessing techniques such as lemmatization and stopword removal are used to refine the textual data further. This ensures that only meaningful terms are retained, providing a standardized and coherent data representation ready for analysis. Additionally, categorical labels in the dataset are converted into numeric form using label encoding, enabling machine learning algorithms to process them effectively. To address class imbalances, upsampling techniques are employed, ensuring equal representation of all classes. Optimizing computational resources by utilizing GPUs and configuring optimizers, also plays a crucial role in enhancing model training efficiency. Finally, multi-dimensional label extraction and model evaluation are conducted to provide comprehensive insights into individual personality dimensions and overall model performance, using metrics such as accuracy, precision, recall, and F1 score.

3.2 Text Representation Techniques

TF-IDF (Term Frequency-Inverse Document Frequency) vectorization converts text into numerical representations by weighting terms based on document frequency and corpus rarity. This enhances classifiers by capturing term importance and aiding in dimensionality reduction, improving predictive accuracy and efficiency in text processing tasks.

GloVe (Global Vectors for Word Representation) embeddings create dense vector representations of words based on their co-occurrence statistics in a corpus. This technique helps classifiers interpret and leverage semantic relationships within text, improving performance in tasks like sentiment analysis and text classification.

BERT (Bidirectional Encoder Representations from Transformers) embeddings produce high-dimensional vectors that capture deep semantic and contextual meanings by processing text bidirectionally. This method enhances classifiers' understanding and prediction accuracy in tasks requiring nuanced text analysis, such as sentiment analysis and classification.

3.3 Machine Learning Models

The diagram in figure 1, illustrates the architecture of the Machine Learning Models used in this research. This crucial step involves various operations such as preprocessing, text representation techniques, Machine Learning models architecture and Performance evaluation matrix.

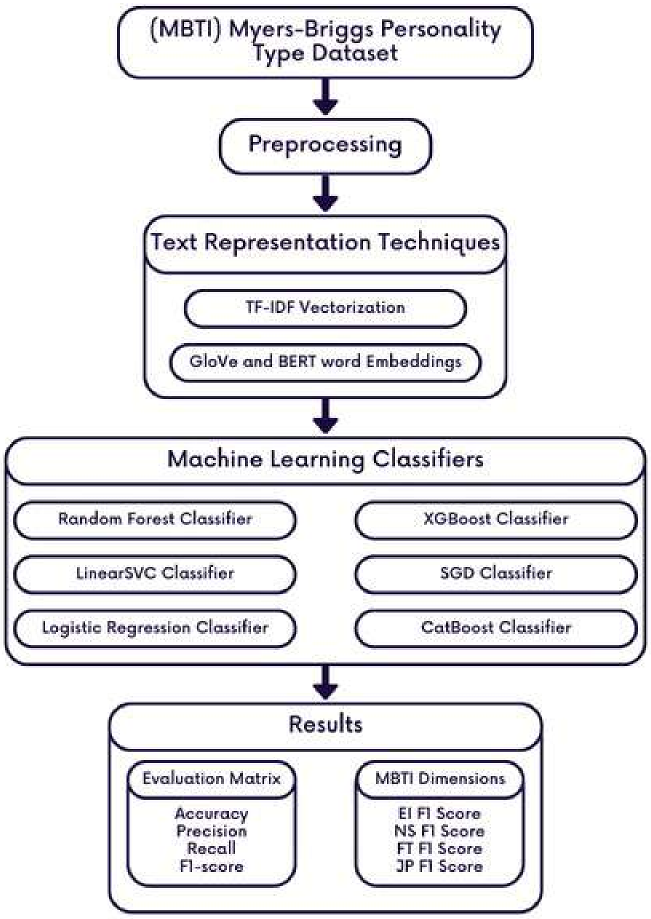

Figure 1. Block diagram of MBTI personality Prediction with word embedding technique

3.3.1 Random Forest Classifier

Random Forest Classifier, an ensemble learning method, constructs multiple decision trees during training and aggregates their outputs for classification. Preprocessing involves data cleaning and standardization, followed by feature extraction using TF-IDF Vectorizer and GloVe or BERT embeddings. The model architecture integrates these components into a pipeline. Training entails considering random feature subsets for each split, mitigating overfitting. Evaluation metrics like accuracy and F1 score assess performance.

3.3.2 XGBoost Classifier

XGBoost, a gradient boosting framework, iteratively builds decision trees to minimize loss during training. Text preprocessing involves standardization, and feature extraction uses TF-IDF Vectorizer and GloVe or BERT embeddings. XGBoost's scalability and efficiency make it suitable for large datasets. Evaluation metrics assess its performance in predicting MBTI personality types.

3.3.3 Linear Support Vector Classifier

LinearSVC, a linear classification model, is efficient in high-dimensional spaces. Text preprocessing includes data cleaning and normalization. GloVe embeddings enrich the model's understanding of text, while TF-IDF Vectorizer converts it into numerical features. LinearSVC is trained on the data, and evaluation metrics like accuracy and F1 score assess its performance.

3.3.4 Stochastic Gradient Descent Classifier

SGD Classifier, known for its efficiency, is combined with GloVe or BERT embeddings for text classification. Text preprocessing involves cleaning and standardization. Feature extraction uses TF-IDF Vectorizer. The model is trained using SGD algorithm, and evaluation metrics assess its performance in predicting MBTI personality types.

3.3.5 Logistic Regression Classifier

Logistic Regression, a linear classification algorithm, is employed with GloVe or BERT embeddings for text classification. Preprocessing includes data cleaning and splitting. Feature extraction uses TF-IDF Vectorizer. The model is trained, and evaluation metrics assess its performance in predicting MBTI personality types.

3.3.6 CatBoost Classifier

CatBoost, a gradient boosting library, is combined with GloVe or BERT embeddings for text classification. Preprocessing involves cleaning and splitting text data. Feature extraction uses TF-IDF Vectorizer. CatBoost is trained with GPU acceleration, and evaluation metrics assess its performance in predicting MBTI personality types.

4 Results and Discussion

This section presents a detailed analysis and interpretation of the findings from the study, focusing on the performance metrics of various classifiers and their effectiveness in predicting MBTI personality dimensions. The discussion integrates these results with existing literature, highlighting the impact of different preprocessing techniques, text representation methods, and the challenges of class imbalance. Through this comprehensive approach, the section aims to provide valuable insights into the study's outcomes, their implications, and potential directions for future research.

4.1 Model Performance

This subsection evaluated the performance of several classifiers applied to the dataset, using key metrics such as accuracy, precision, recall, and F1-score. Additionally, the performance of each classifier was assessed across the four MBTI dimensions: Extraversion-Introversion (EI), Intuition-Sensing (NS), Thinking-Feeling (FT), and Judging-Perceiving (JP).

4.1.1 Overall Classifier Performance

The table below summarizes the performance of the classifiers:

4.1.2 MBTI Dimension-Specific Performance

The table below provides the performance metrics for each classifier across the four MBTI dimensions:

Table 3. MBTI Dimensions for F1-Score Performance of Machine Learning Classifiers with GloVe embeddings

Classifiers	EI	NS	FT	JP
Random Forest	98.85%	99.54%	98.57%	98.03%
XGBoost	98.82%	99.40%	98.68%	98.65%
LinearSVC	98.05%	99.12%	97.63%	96.93%
SGD	92.41%	92.41%	93.26%	92.29%
Logistic Regression	96.27%	97.83%	96.20%	95.65%
CatBoost	94.11%	95.91%	93.97%	93.37%

Table 4. MBTI Dimensions for F1-Score Performance of Machine Learning Classifiers with BERT embeddings

Classifiers	EI	NS	FT	JP
Random Forest	98.98%	99.75%	98.18%	98.23%
XGBoost	97.36%	98.28%	96.94%	96.68%
LinearSVC	98.13%	98.97%	97.37%	97.01%
SGD	95.88%	96.66%	95.69%	94.95%
Logistic Regression	96.64%	97.29%	96.61%	95.23%
CatBoost	94.89%	96.36%	94.98%	94.00%

The results from the analysis revealed that XGBoost was the top-performing model, particularly with GloVe embeddings, achieving the highest overall accuracy of 97.26%. However, Random Forest demonstrated consistent performance across both GloVe and BERT embeddings, indicating its robustness in personality type classification. The dimension-specific analysis further highlighted the classifiers' strengths in capturing nuanced personality traits, especially in the NS and EI dimensions. When compared to existing literature, the findings aligned with the notion that ensemble methods like XGBoost and Random Forest excel in complex classification tasks. However, the variation in performance across different MBTI dimensions suggested that certain traits were more challenging to predict, requiring further exploration and potentially more sophisticated models.

4.2 Discussion

The discussion section interprets the results obtained from the analysis, highlighting the effectiveness of different classifiers in predicting MBTI personality types. XGBoost emerged as the top-performing model, followed by Random Forest and LinearSVC. The dimension-specific analysis underscores the classifiers' strengths in capturing nuanced personality traits. Additionally, the section compares the results with existing literature, discusses implications, and addresses limitations and challenges encountered during the study. Overall, the study contributes to the understanding of machine learning in personality type classification and suggests avenues for future research.

Table 1. Performance Evaluation of Overall Machine Learning Classifiers with GloVe embeddings

Classifiers	Accuracy	Precision	Recall	F1-Score
Random Forest	96.86%	96.77%	96.79%	96.78%
XGBoost	97.26%	97.24%	97.26%	97.24%
LinearSVC	95.10%	94.97%	95.03%	94.98%
SGD	85.48%	85.41%	85.38%	84.78%
Logistic Regression	91.93%	91.72%	91.86%	91.75%
CatBoost	87.24%	87.12%	87.47%	87.21%

Table 2. Performance Evaluation of Overall Machine Learning Classifiers with BERT embeddings

Classifiers	Accuracy	Precision	Recall	F1-Score
Random Forest	96.86%	96.75%	96.74%	96.73%
XGBoost	93.96%	93.81%	93.86%	93.79%
LinearSVC	94.93%	94.92%	95.01%	94.93%
SGD	90.53%	90.89%	90.16%	89.92%
Logistic Regression	91.31%	91.07%	91.42%	91.16%
CatBoost	88.87%	88.60%	88.81%	88.65%

4.3 Analysis of Preprocessing Techniques

The impact of preprocessing techniques, including text cleaning, tokenization, TF-IDF vectorization, and word embeddings (GloVe and BERT), was critical in optimizing classifier performance. The results demonstrated that these techniques significantly enhanced the quality of textual data, leading to more accurate predictions. The comparison of GloVe and BERT embeddings indicated that while both were effective, BERT generally provided more consistent results across classifiers.

4.4 Class Imbalance and Upsampling

Class imbalance posed a significant challenge in the dataset, potentially skewing the performance of classifiers. The implementation of upsampling techniques effectively mitigated this issue, as evidenced by improved performance metrics across the board. The analysis underscored the importance of addressing class distribution in classification tasks to ensure fair and accurate predictions.

4.5 Discussion of Results

This section synthesized the study's findings, offering a comprehensive analysis of model performance, MBTI dimension-specific results, and the implications of these outcomes. The discussion also addressed the limitations encountered, such as challenges with specific classifiers and the computational complexity of certain models. Moreover, the ethical considerations in applying machine learning to personality prediction were explored, emphasizing the need for transparency and fairness in model development.

5 Conclusion

In conclusion, this study provided a thorough examination of various machine learning classifiers in predicting MBTI personality types based on textual data. XGBoost emerged as the top performer, followed by Random Forest and LinearSVC. The analysis of MBTI dimensions highlighted the classifiers' capabilities in capturing nuanced personality traits, particularly in the NS and EI dimensions. Despite challenges with certain classifiers, the research demonstrated the importance of model selection and preprocessing techniques in achieving accurate predictions.

The findings have significant implications for fields such as human resources and personalized recommendation systems, where understanding personality traits can enhance decision-making processes. Future research should focus on exploring ensemble methods, neural network architectures, and interpretability techniques to further improve the robustness and applicability of personality prediction models in real-world settings. This study contributed valuable insights to the advancement of machine learning approaches for personality prediction and set the stage for future exploration in this dynamic field.

Author Contributions

Nimra Ashraf: Conceptualisation, data collection, supervision, methodology.**Rao Sohail Iqbal**: Data collection, experimental work, and writing the original draft.**Sehar Bano**: Data labelling, proofreading.**Hafiz Muhammad Azeem**: Algorithm testing, review of the original draft.**Shagufta Naz**:Data collection, validation, editing.

Compliance with Ethical Standards

It is declare that all authors don't have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors. Furthermore, informed consent was obtained from all individual participants included in the study.

References

[1] S. J. Johnson, M. R. Murty, *et al.*, "Machine learning approach to improve data connectivity in text-based personality prediction using multiple data sources mapping," *Journal of Scientific & Industrial Research*, vol. 82, no. 1, pp. 109–119, 2023.

[2] T. Wang *et al.*, "Modeling digital personality: A fuzzy-logic-based myers-briggs type indicator for fine-grained analytics of digital human," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 1, pp. 1096–1107, 2023.

[3] D. Radisavljević *et al.*, "Myers-briggs type indicator and the big five model-how our personality affects language use," in *2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pp. 1–6, IEEE, 2022.

[4] Y. Justindhas, S. M. Mohanraj, and R. Shivani, "A synoptic survey on personality prediction system using mbti," in *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 950–955, IEEE, 2022.

[5] G. Praphulla *et al.*, "Personality prediction using machine learning techniques," in *AIP Conference Proceedings*, vol. 2794, AIP Publishing, 2023.

[6] R. Hegde *et al.*, "Personality classification using data mining approach," *IJRAR (International Journal of Research and Analytical Reviews)*, vol. 6, no. 1, pp. 1–6, 2019.

[7] N. Qamar and A. A. Malik, "A quantitative assessment of the impact of homogeneity in personality traits on software quality and team productivity," *IEEE Access*, vol. 10, pp. 122092–122111, 2022.

[8] M. Sharma, S. Joshi, S. Sharma, A. Singh, and R. Gupta, "Data mining classification techniques to assign individual personality type and predict job profile," in *Proc. 2021 9th Int. Conf. Reliability, Infocom Technol. Optimization (Trends Future Directions) (ICRITO)*, pp. 1–5, 2021.

[9] D. Vasundhara *et al.*, "Mbti (myers briggs type indicator) based personality classification," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 6, pp. 4652–4657, 2023.

[10] G. Ryan, P. Katarina, and D. Suhartono, "Mbti personality prediction using machine learning and smote for balancing data based on statement sentences," *Information*, vol. 14, no. 4, p. 217, 2023.

[11] Z. Wang *et al.*, "Smotetomek-based resampling for personality recognition," *IEEE Access*, vol. 7, pp. 129678–129689, 2019.

[12] S. S. Keh and I. Cheng, "Myers-briggs personality classification and personality-specific language generation using pre-trained language models," *arXiv preprint arXiv:1907.06333*, 2019.

[13] M. H. Amirhosseini and H. Kazemian, "Machine learning approach to personality type prediction based on the myers-briggs type indicator," *Multimodal Technologies and Interaction*, vol. 4, no. 9, 2020.

[14] A. Talasbek *et al.*, "Personality classification experiment by applying k-means clustering," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 16, pp. 162–177, 2020.

[15] C. Zhao *et al.*, "Relationship between personality types in mbti and dream structure variables," *Frontiers in Psychology*, vol. 11, 2020.

[16] K. Sönmezöz, Uğur, and B. Diri, "Mbti personality prediction with machine learning," in *Proc. 2020 28th Signal Process. Commun. Appl. Conf. (SIU)*, pp. 1–4, 2020.

[17] S. Patel *et al.*, "Personality analysis using social media," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 3, pp. 306–309, 2020.

[18] V. Ramachandran *et al.*, "Myers-briggs type indicator in medical education: A narrative review and analysis," *Health Professions Education*, vol. 6, no. 1, pp. 31–46, 2020.

[19] A. S. Khan *et al.*, "Personality classification from online text using machine learning approach," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, pp. 460–476, 2020.

[20] M. N. Sahono, F. U. Sidiastahta, G. F. Shidik, A. Z. Fanani, S. Nuraisha, and E. Lutfina, "Extrovert and introvert classification based on myers-briggs type indicator (mbti) using support vector machine (svm)," in *Proc. 2020 Int. Seminar Appl. Technol. Inf. Commun. (iSemantic)*, pp. 572–577, 2020.

[21] N. Haziqah *et al.*, "Improving intelligent personality prediction using myers-briggs type indicator and random forest classifier," 2020.

[22] S. Choi, "The interdependency of the diction and mbti personality type of online users," *American Journal of Applied Psychology*, vol. 10, no. 1, pp. 21–26, 2021.

[23] C. Basto, "Extending the abstraction of personality types based on mbti with machine learning and natural language processing," *arXiv preprint arXiv:2105.11798*, 2021.

[24] R. L. Vásquez and J. Ochoa-Luna, "Transformer-based approaches for personality detection using the mbti model," in *2021 XLVII Latin American Computing Conference (CLEI)*, pp. 1–7, IEEE, 2021.

[25] H. Shafi, "A machine learning approach for personality type identification using mbti framework," *Journal of Independent Studies and Research: Computing*, vol. 19, no. 2, 2021.

[26] S. Ontoum and J. H. Chan, "Personality type based on myers-briggs type indicator with text posting style by using traditional and deep learning," *arXiv preprint arXiv:2201.08717*, 2022.

[27] Y. Zhao *et al.*, "Reading personality preferences from motion patterns in computer mouse operations," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1619–1636, 2020.

[28] D. Russo and K. J. Stol, "Gender differences in personality traits of software engineers," *IEEE Transactions on Software Engineering*, vol. 48, no. 3, pp. 819–834, 2020.

[29] H. Naik *et al.*, "Myers briggs type indicator (mbti)-personality prediction using deep learning," in *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*, pp. 1–6, IEEE, 2022.

[30] Y. Zhou, J. Shi, and Q. Yu, "Mbti personality analysis and prediction," 2022.

[31] V. G. D. Santos and I. Paraboni, "Myers-briggs personality classification from social media text using pre-trained language models," *arXiv preprint arXiv:2207.04476*, 2022.

[32] P. K. N. and M. L. Gavrilova, "Latent personality traits assessment from social network activity using contextual language embedding," *IEEE Transactions on Computational Social Systems*, vol. 9, no. 2, pp. 638–649, 2021.

[33] N. Kovačević *et al.*, "Personality trait recognition based on smartphone typing characteristics in the wild," *IEEE Transactions on Affective Computing*, vol. 14, no. 4, pp. 3207–3217, 2023.

[34] N. Soonpipatskul *et al.*, "Personality perceptions of conversational agents: A task-based analysis using thai as the conversational language," *IEEE Access*, 2023.

[35] M. Pessoa *et al.*, "A journey to identify users' classification strategies to customize game-based and gamified learning environments," *IEEE Transactions on Learning Technologies*, 2023.