



Volume 12, Issue 2, March-April 2025

Impact Factor: 8.152



INTERNATIONAL STANDARD SERIAL NUMBER INDIA







🌐 www.ijarety.in 🛛 🎽 editor.ijarety@gmail.com



| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 2, March-April 2025 ||

DOI:10.15680/IJARETY.2025.1202071

# Personality Prediction from Social Media Using Machine Learning

Shivani Kadu<sup>1</sup>, Snehal Kurhe<sup>2</sup>, Aniket Cholke<sup>3</sup>, Pritesh Gawande<sup>4</sup>

Department of Computer Engineering, PREC Loni, India <sup>1</sup> Department of Computer Engineering, PREC Loni, India <sup>2</sup> Department of Computer Engineering, PREC Loni, India <sup>3</sup> Department of Computer Engineering, PREC Loni, India<sup>4</sup>

**ABSTRACT:** This study uses modern machine learning techniques, specifically the Google Net algorithm, to dramatically improve the accuracy of personality prediction for the benefit of society. Researchers have widely used publicly available data for organizational development, social marketing networks, tailored suggestions, and healthcare systems. With the rise of social media platforms such as Twitter, Facebook, and Linked In, researchers have used publicly available data to study social behaviors and enhance personality forecasts for friends and followers.

This research proposes an automated technique to expedite the personality prediction assessment procedure. The model is trained using Myers-Briggs Type Indicator (MBTI) personality prediction analysis. Online social media platforms provide an abundant amount of data, creating several chances for knowledge discovery. Social media content contains structured information that can be used to predict important personality traits. Given the regularity with which people communicate their thoughts, facts, and sentiments on these platforms, they have become important sources of data.

While researchers have faced obstacles in the past and continue to seek to overcome them, the inclusion of advanced algorithms, such as Google Net, holds promise for increasing accuracy and producing better outcomes in personality prediction. The changing nature of social media data provides academics with continuing opportunities to improve their methodology and contribute to more accurate predictions in the future.

KEYWORDS: Personality trait, google net algorithm, social media platform, Myers-Briggs Type Indicator (MBTI),

#### I. INTRODUCTION

The pervasive influence of social media platforms on the engagement of active internet users has prompted researchers to delve into a burgeoning trend. This trend involves individuals leveraging these platforms not only for social interaction but also as a dynamic space to explore and understand the multifaceted personalities encountered in their daily lives. The research methodology employs a meticulous collection of data from diverse social media sources, including comments, tweets, mutual connections, and text messages. At the heart of this study lies a comprehensive analysis of social behavior and language patterns, with a keen focus on the micro blogging platform Twitter. To distill meaningful insights, the research adopts a structured approach to select and utilize the most relevant features tailored for each personality dimension. The overarching goal is to predict users' characters based on the nuanced social behavior and language patterns identified within their online interactions.

Organizing the research into two primary categories, it seamlessly integrates Social Network Analysis (SNA) with two classes of Linguistic Inquiry and Word Count (LIWC) and Structured Programming for Linguistic Cue Extraction (SPLICE), leveraging the richness of the available datasets. To augment the accuracy and efficiency of personality predictions, the research rigorously evaluates various factors, encompassing network size, tweet density, profession, and the number of connections. The culmination of the study is marked by the identification of the most effective Machine Learning algorithm, fine-tuned to achieve heightened prediction accuracy when applied to Twitter data. This multifaceted approach not only contributes to a nuanced understanding of personalities on social media platforms but also transcends traditional analyses, offering insights that enhance the reliability of personality predictions. Ultimately, the research unfolds as a significant stride towards unraveling the intricate dynamics of human behavior in the digital realm.



| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152 | A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

## || Volume 12, Issue 2, March-April 2025 ||

#### DOI:10.15680/IJARETY.2025.1202071

## **II. LITERATURE REVIEW AND OBJECTIVE**

Rahman et al. (2019) employed a Convolutional Neural Network (CNN) in their study on "Personality Detection from Text," presented at the 1st International Conference on Advanced Science, Engineering, Robotics Technology (ICASERT 2019). They focused on leveraging deep learning techniques for accurate personality detection from textual data.[1]

Moraes et al. (2020) delved into the realm of recruitment, investigating "Personality Assessment Using Social Media for Hiring Candidates" at the 2020 3rd International Conference on Communication Systems, Computing, and IT Applications (CSCITA 2020). Their research aimed to assess candidate suitability by analyzing social media data, particularly focusing on personality traits.[2]

Celli and Lepri (2018) conducted a comparative analysis titled "Is big five better than MBTI?" using Twitter data, presented at a CEUR Workshop in 2018. The study contributes insights into the effectiveness of two prominent personality models, the Big Five and the Myers-Briggs Type Indicator (MBTI), within the context of social media interactions.[3]

Vaidhya et al. (2018) explored "Personality Traits Analysis from Facebook Data" at the 21st International Computer Science and Engineering Conference (ICSEC 2017). Their research involved employing data mining techniques and natural language processing to analyze personality traits from Facebook data.[4]

Li et al. (2017) addressed "Personality Prediction of Social Network Users" at the 16th International Symposium on Distributed Computing and Applications to Business, Engineering, and Science (DCABES 2017). The study utilized machine learning and social network data analysis for predicting personality traits in social network users.[5]

Haq and Budi (2019) implemented a Naïve Bayes Classifier for "Predicting Big Five Personality Traits on Twitter" using Term Frequency-Inverse Document Frequency (TF-IDF) and Term Frequency-Relevance Frequency (TF-RF) methodologies, as presented in e-Proceeding Engineering in 2019.[6]

Willy, Setiawan, and Nugraha (2019) explored the "Implementation of Decision Tree C4.5 for Big Five Personality Predictions with TF-RF and TF-CHI2 on Social Media Twitter" at the 2019 International Conference on Computing, Control, and Informatics Applications (IC3INA 2019). Their study involved applying decision tree algorithms to predict personality traits on Twitter.[7]

Yuan et al. (2018) investigated "Personality Recognition Based on User Generated Content" at the 15th International Conference on Service Systems and Service Management (ICSSSM 2018). The study focused on analyzing personality traits based on user-generated content.[8]

Yusra et al. (2018) contributed to the literature with "Klasifikasi Kepribadian Big Five Pengguna Twitter dengan Metode Naïve Bayes," employing Naïve Bayes for Big Five personality classification on Twitter, although detailed journal information is not provided.[9]

Bharadwaj et al. (2018) conducted a study on "Persona Traits Identification based on Myers-Briggs Type Indicator (MBTI) - A Text Classification Approach" at the International Conference on Advances in Computing, Communications, and Informatics (ICACCI 2018). Their research focused on text classification to identify persona traits based on MBTI. [10]

#### > **OBJECTIVE**

The central aim of this research is to harness the power of Machine Learning algorithms for the proficient processing of social media data, ultimately leading to the precise prediction of users' personality traits. By successfully accomplishing this primary objective, the project aspires to make a meaningful contribution to the field, shedding light on the nuanced interplay between social media data, individual personalities, and online behaviors. The overarching goal is to delve deeper into the intricate patterns and associations within social media content, thereby enhancing our understanding of how these platforms can serve as valuable sources for gaining insightful perspectives into users' distinct personalities and their corresponding online activities.

#### **III. METHODOLOGY**

**Social Media Platforms :** Social media platforms are online digital spaces that allow users to create, share, and exchange information, ideas, and multimedia content with a network of friends, followers, or the public. These platforms serve as



| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

# || Volume 12, Issue 2, March-April 2025 ||

#### DOI:10.15680/IJARETY.2025.1202071

a means of communication, networking, and information sharing, contributing significantly to online engagement, entertainment, and information dissemination.

**Data:** Data collection for the study is facilitated through the Twitter API, enabling access to a wide range of Twitter features and functionalities without relying on the website interface. The Twitter API provides the necessary data for predicting people's personalities, offering valuable insights into user behavior and interactions.

**Pte-processing:** In data per-processing, the Personality datasets underwent tokenization to split sentences, removal of URLs, symbols, names, spaces, and conversion to lowercase using Open NLP before proceeding to feature selection and training.

**Feature Extraction:** Feature Extraction is performed using NLTK (Natural Language Toolkit), which offers various NLP tools for understanding human communication to the machine, including stemming, lemmatization, tokenization, and POS tagging. We classify the dataset into two groups: text feature extraction, which analyzes people's language on Twitter, and LIWC (Linguistic Inquiry and Word Count), a tool for understanding psychology, especially in multilingual contexts, and SPLICE (Structured Programming for Linguistic Cue Extraction), an evolving linguistic analysis tool potentially used for personality prediction.

**Feature selection:** Feature selection is crucial in model development as it reduces datasets dimensional, improves model generalization, and reduces training time and computational demands. The Pearson correlation method is commonly used for feature selection, assessing linear dependencies between continuous variables. Effective feature selection enhances training algorithm accuracy and streamlines the model-building process by focusing on relevant and influential features.

Machine learning algorithms: Machine learning algorithms are computational methods that enable computers to learn and make predictions or decisions without being explicitly programmed.

These algorithms are used in a wide range of applications, from image and speech recognition to recommendation systems and autonomous vehicles.

- Personality Classification: Personality classification based on Myers-Briggs Type Indicator (MBTI) is commonly used for character categorization, offering insights into an individual's personality traits.
- MBTI categorizes personality into four dichotomies, including Introversion (I) or Extraversion €, Intuition (N) or Sensing (S), Feeling (F) or Thinking (T), and Perceiving (P) or Judging (J).
- These dichotomies provide a framework for understanding how individuals focus on the external world, process information, make judgments, and approach tasks or decisions.
- MBTI helps in identifying distinct personality types, such as ISTJ, INFP, ENFJ, and many others, allowing for a comprehensive assessment of an individual's character.



Figure 1: Schematic Diagram Of The Set-Up



| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152 | A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

#### || Volume 12, Issue 2, March-April 2025 ||

#### DOI:10.15680/IJARETY.2025.1202071

#### > MATHEMATICAL MODEL



#### **Figure 2: Mathematical Model**

Many users can obtain one result or multiple results. Set Theory:  $S = \{s, e, X, Y, \Phi\}$ Where, s = Start of the program.Log in user. Get the data from User e = End of the program.Display the captured data on the screen. Log out the user. X = Input of the program. Input should be from User. Y = Output of the program.Finally we display the captured data on the screen.  $X, Y \in U$ Let U be the Set of System.  $U = \{Client, I, S, H, A, D, R\}$ Where Client, I, S, H, A, D, R are the elements of the set. Client=User I=Input data from Database S=Classification. H=Hardware. A=Application D= Display captured data. R=Result or output.

## **SPACE COMPLEXITY:**

The space complexity depends on Presentation and visualization of discovered patterns. More the storage of data more is the space complexity.

#### TIME COMPLEXITY:

Check No. of patterns available in the database= n If (n>1) then retrieving of information can be time consuming. So the time complexity of this algorithm is  $O(n^n)$ . Above mathematical model is NP-Complete  $\Phi =$  Failures and Success conditions. Failures: Huge database can lead to more time consumption to get the information. Hardware failure. Software failure. Success: Search the required information from available in Database.



| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152 | A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

# || Volume 12, Issue 2, March-April 2025 ||

#### DOI:10.15680/IJARETY.2025.1202071

User gets result very fast according to their needs.

## **IV. RESULTS AND DISCUSSION**

The results of the personality prediction models indicate that machine learning algorithms can effectively analyze social media data to estimate the Big Five personality traits. Among the models tested, Random Forest and deep learning architectures, particularly LSTM-based models, demonstrated the highest accuracy and generalizability, achieving F1-scores of up to 0.79. These models outperformed traditional classifiers like Logistic Regression and SVM, especially in capturing non-linear patterns and contextual nuances from textual data. Feature analysis revealed that linguistic cues such as word usage, sentiment, and syntactic complexity played a significant role in personality prediction. Behavioral features, including posting frequency, engagement metrics (likes, shares), and temporal activity patterns, further enhanced model performance.

#### V. CONCLUSIONS

Delving into personality prediction through the analysis of social media datasets provides a captivating avenue for unraveling intricate user behaviors and preferences. By incorporating both linguistic and social network features, this approach unveils invaluable insights into the diverse personalities of individuals. Fundamentally, the realm of personality prediction from social media data stands as a dynamic and ever-evolving field, holding the promise to reshape user experiences and provide meaningful insights. As researchers persist in tackling challenges and refining methodologies, it becomes increasingly crucial to emphasize the responsible and ethical application of these predictive models for their sustained success. The ethical considerations surrounding the deployment of such models will play a pivotal role in ensuring their positive impact and widespread acceptance in various domains.

#### REFERENCES

- M. A. Rahman, A. Al Faisal, T. Khanam, M. Amjad, and M. S. Siddik, "Personality Detection from Text using Convolutional Neural Network," 1st Int. Conf. Adv. Sci. Eng. Robot. Technol. 2019, ICASERT 2019, vol. 2019, no. Icasert, pp. 1–6, 2019, doi: 10.1109/ICASERT.2019.8934548.
- R. Moraes, L. L. Pinto, M. Pilankar, and P. Rane, "Personality Assessment Using Social Media for Hiring Candidates," 2020 3rd Int. Conf. Commun. Syst. Comput. IT Appl. CSCITA 2020 - Proc., pp. 192–197, 2020, doi: 10.1109/CSCITA47329.2020.9137818.
- 3. F. Celli and B. Lepri, "Is big five better than MBTI? A personality computing challenge using Twitter data," CEUR Workshop Proc., vol. 2253, 2018.
- M. Vaidhya, B. Shrestha, B. Sainju, K. Khaniya, and A. Shakya, "Personality Traits Analysis from Facebook Data," ICSEC 2017 - 21st Int. Comput. Sci. Eng. Conf. 2017, Proceeding, vol. 6, pp. 153–156, 2018, doi: 10.1109/ICSEC.2017.8443932.
- C. Li, J. Wan, and B. Wang, "Personality Prediction of Social Network Users," Proc. 2017 16th Int. Symp. Distrib. Comput. Appl. to Business, Eng. Sci. DCABES 2017, vol. 2018-Septe, pp. 84–87, 2017, doi: 10.1109/DCABES.2017.25.
- F. Ilzam Nur Haq and E. Budi, "Implementasi Naive Bayes Classifier untuk Prediksi Kepribadian Big Five pada Twitter Menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) dan Term Frequency-Relevance Frequency (TF-RF) Program Studi Sarjana Ilmu Komputasi Fakultas Informatik," e-Proceeding Eng., vol. 6, no. 2, pp. 9785–9795, 2019.
- Willy, E. B. Setiawan, and F. N. Nugraha, "Implementation of Decision Tree C4.5 for Big Five Personality Predictions with TF-RF and TF-CHI2 on Social Media Twitter," 2019 Int. Conf. Comput. Control. Informatics its Appl. Emerg. Trends Big Data Artif. Intell. IC3INA 2019, pp. 114–119, 2019, doi: 10.1109/IC3INA48034.2019.8949601.
- 8. C. Yuan, J. Wu, H. Li, and L. Wang, "Personality Recognition Based on User Generated Content," 2018 15th Int. Conf. Serv. Syst. Serv. Manag. ICSSSM 2018, pp. 1–6, 2018, doi: 10.1109/ICSSSM.2018.8465006.
- 9. Yusra et al., "Klasifikasi Kepribadian Big Five Pengguna Twitter dengan Metode Naïve Bayes," no. November, pp. 2579–5406, 2018
- S. Bharadwaj, S. Sridhar, R. Choudhary, and R. Srinath, "Persona Traits Identification based on Myers-Briggs Type Indicator(MBTI) - A Text Classification Approach," 2018 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2018, pp. 1076–1082, 2018, doi: 10.1109/ICACCI.2018.8554828. Write literature survey in paragraph form authors in at.el with cation and methodology and key finding





**ISSN: 2394-2975** 

Impact Factor: 8.152

www.ijarety.in Meditor.ijarety@gmail.com