International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

"Personality Prediction Using ML And NLP"

1. Prof. Afsha Akkalkot,2. Abhishek Nagarkar,3.Shrikant Nagawade,4.Rahul Rale,5. Suraj Raut

[1]*Assistant Professor, Department of Computer Engineering, Zeal college of engineering and research narhe pune.*

[2,3,4,5]*Students, Department of Computer Engineering, Zeal college of engineering and research narhe Pune*

Abstract: ***Predicting an individual's personality can help to understand different aspects related to them such as how do they deal with stress, openness to social activities, and if they influence others. One of the key characteristics that affects how people interact with the outside environment showcases their personality. Personality prediction systems have gained significant attention due to their potential applications in various domains, including psychology, marketing, job recruitments, and human resources. Traditional approaches to personality prediction primarily rely on manual evaluation of self-reported questionnaire, which suffers from several limitations. The limitations include subjectivity biases, potential social desirability effects, and time-consuming data collection processes. This project aims to leverage Machine Learning (ML) and Natural Language Processing (NLP) techniques to predict personality traits of an individual using questionnaire responses. Personality prediction system has been implemented based on Myers-Briggs Type Indicator (MBTI) model which is a popular personality prediction framework. The model consists of 16 personality types based on combination of 4-pairs of personality traits.***

Keywords: Personality Prediction, Machine Learning, Natural Language Processing, Myers-Briggs Type Indicator.

I. INTRODUCTION

Personality refers to a unique characteristics and traits of a person based on their thoughts, feelings and behaviours that distinguish a person from others. It's been done for a long time to determine a person's personality based on their nature. Traditional approaches for personality prediction were based on the responses received for a self-reported questionnaire. This way of personality prediction was time-consuming and depends on manual inspection which can lead to human error. In recent years, there has been a significant growth in the field of personality prediction using Machine Learning and Natural Language Processing techniques.

We have used and implemented a popular model for personality prediction which is Myers-Briggs Type Indicator (MBTI), which categorizes individuals into distinct personality types based on their preferences in four key dimensions:

1. Extraversion / Introversion
2. Sensing / Intuition
3. Thinking / Feeling
4. Judging / Perceiving

1.1 PROBLEM STATEMENT

Personality prediction is the task of inferring a person's personality traits from their responses to questionnaire. Personality is a complex trait that is difficult to measure. Traditional approaches to personality prediction were time consuming and cumbersome and might lead to biased predictions. The process was

rigorous and needed to be done manually. The proposed system provides a solution for Personality Prediction of a person using Machine Learning and Natural Language Processing techniques and algorithms. This system will provide the organizations with a better, faster and more accurate prediction of personality without any need of expert person.

II.RELATED WORKS

R. Hegde, S. Hegde, Sanjana, S. Kotian, and S. Shetty, [1] describes, “Personality classification using Data mining approach”. In this paper, the author said that Personality classification refers to the psychological classification of different types of individuals. This project deals with the areas where it determines the characteristics of a person. It can be helpful to classify person using Personality classification using data mining approach. In this paper, we aim to automate the personality prediction of the users by taking a personality test. The system uses classification algorithm i.e., N-closest neighbourhood algorithm (NCN). The analysis is done using vast set of data in data set and has been compared with the user input. This paper mainly focuses on classification algorithm.

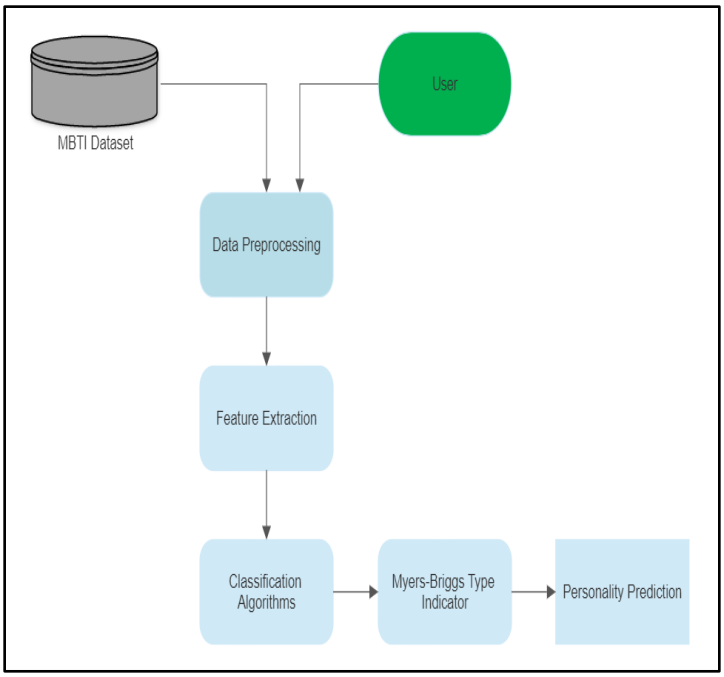
F. Q. Annisa, E. Supriyanto and S. Taheri, [2] describes, "Personality Dimensions Classification with EEG Analysis using Support Vector Machine”. In this paper, the author said that Personality is the fundamental thing that forms the behavioural tendencies of each individuality in a situation. A common model used to describe personality is the big five personality that divides personality traits into five dimensions of neuroticism, extraversion, openness, agreeableness, and conscientiousness. Personality assessment through physiological signals offers objectivity and reliability of the test results due to the minimal role of test takers in the examination process. One widely recommended approach is signal-based analysis of electroencephalography (EEG). The EEG signal feature of the ASCERTAIN public database was extracted using discrete wavelet transform (DWT) and was classified using support vector machine (SVM) to determine personality dimensions. The results showed better performance compared to the application of other techniques on the same dataset with 69% and 75.9% accuracy to determine extraversion and neuroticism level, respectively. However, this accuracy still needs to be improved to generate reliable model. Increased data variability can be useful for understanding brain dynamic activity per individual.

D. Al-Hammadi and R. K. Moore, [3] describes, "Using Sampling Techniques and Machine Learning Algorithms to Improve Big Five Personality Traits Recognition from Non-verbal Cues". In this paper, the author said that Automatic personality recognition using the Big Five dimensions (OCEAN: extraversion, agreeableness, conscientiousness, neuroticism and openness) is capturing the attention of researchers. Personality recognition is expected to have encouraging future in Human- computer and Robot Interaction applications. Human speech conveys rich information that can be derived to recognize speaker traits. However, our focus is on the rich content of non-verbal features in human speech. We focus on how humans talk, not what they talk about. The focus in this paper is to experiment with four different machine learning techniques, and their performance in recognizing personality traits, we report our results in this regard. We use the Speaker Personality Corpus provided by the Interspeech 2012 challenge. First, we recognize three issues affecting the system's low performance: dimensionality, judges' agreement, and imbalanced data. Next, we address each issue and provide a solution to improve the system's performance. Finally, we compare our results with the baseline showing better classification results.

Alam Sher Khan, Hussain Ahmad, Muhammad Zubair Asghar, Furqan Khan Saddozai, Areeba Arif and Hassan Ali Khalid, [4] describes, "Personality Classification from Online Text using Machine Learning Approach ". In this paper, the author said that Personality refers to the distinctive set of characteristics of a person that effect their habits, behaviour's, attitude and pattern of thoughts. Text available on Social Networking sites provide an opportunity to recognize individual's personality traits automatically. In this proposed work, Machine Learning Technique, XGBoost classifier is used to predict four personality traits based on Myers-Briggs Type Indicator (MBTI) model, namely Introversion-Extroversion(I-E), Ntuition-Sensing(N-S), Feeling-Thinking(F-T) and Judging-Perceiving(J-P) from input text. Publically available benchmark dataset from Kaggle is used in experiments. The skewness of the dataset is the main issue associated with the prior work, which is minimized by applying Re-sampling technique namely random over-sampling, resulting in better performance. For more exploration of the personality from text, pre-processing techniques including tokenization, word stemming, stop words elimination and feature selection using TF IDF are also exploited. This work provides the basis for developing a personality identification system which could assist organization for recruiting and selecting appropriate personnel and to improve their business by knowing the personality and preferences of their customers. The results obtained by all classifiers across all personality traits is good enough, however, the performance of XGBoost classifier is outstanding by achieving more than 99% precision and accuracy for different traits.

Z. Mushtaq, S. Ashraf and N. Sabahat, [5] describes, "Predicting MBTI Personality type with K-means Clustering and Gradient Boosting". In this paper, the author said that Personality refers to a characteristic pattern of thoughts, behavior, and feelings that makes a person unique. Asking users to fill a questionnaire to get their personality insights could be inaccurate because the users are conscious and try to take a careful approach when filling the survey. However, when it comes to social media, users do not take any consideration before posting their opinions on social media. Therefore, the data obtained from social media could be precious to determine the user personality type. In this paper, we propose a way to analyze the user's data posted on social media by combining two existing machine learning algorithms, such as K-Means Clustering and Gradient Boosting, in order to predict user personality type. Moreover, this research helps to analyze the empirical relation between the user's data posted on social media and the user's personality. In this paper, we used The Myer-Briggs Type Indicator (MBTI) introduced by Swiss psychiatrist Carl Jung. MBTI is based on sixteen personality types, and they act as a valuable reference point to understand a person's unique personality. The technique of combining these two machine learning algorithms gave accurate results than the traditional naive Bayes classification and other algorithms. Results of this study can help bloggers and social media users to know what type of personality they are showing on the social media with the data they posted on the internet.

III.PROPOSED SYSTEM



The system we want to develop is based on MBTI (Myers-Briggs Type Indicator) Model. We are using Machine Learning and Natural Language Processing for the purpose of personality prediction. We have gathered the dataset from Kaggle for personality prediction. We are going to predict the personality through questionnaire which contains objective and subjective questions. We are implementing multiple Machine Learning models and compare them for better accuracy.

The Myers-Briggs Type Indicator (MBTI) is a psychological framework and assessment tool used to categorize individuals into different personality types. It was developed by Katharine Cook Briggs and her daughter Isabel Briggs Myers, based on the theories proposed by Carl Jung.

The MBTI model classifies people into 16 distinct personality types, each represented by a four-letter code. The four dimensions of personality used in the MBTI are:

1. Extraversion (E) vs. Introversion (I):

This dimension reflects how individuals gain energy and focus their attention. Extraverts tend to be outgoing, social, and energized by interacting with others, while introverts prefer solitude, introspection, and gain energy from being alone.

2. Sensing (S) vs. Intuition (N):

This dimension describes how individuals perceive and gather information. Sensors rely on their five senses and prefer concrete, factual information, while intuitive individuals focus on patterns, possibilities, and abstract concepts.

3. Thinking (T) vs. Feeling (F):

This dimension refers to how individuals make decisions and evaluate information. Thinkers tend to rely on logical analysis and objective criteria, while feelers prioritize personal values, empathy, and consider the impact on others.

4. Judging (J) vs. Perceiving (P):

This dimension reflects how individuals prefer to approach the outside world. Judgers prefer structure, organization, and decisiveness, while perceivers are more flexible, adaptable, and open to new experiences.

Combining these four dimensions creates 16 possible personality types, such as ISTJ, ENFP, ENTJ, etc. Each type represents a unique combination of preferences and is associated with different strengths, communication styles, and ways of approaching tasks and relationships.

The MBTI is commonly used in career counselling, team-building exercises, and personal development to gain insights into individuals' preferences, strengths, and potential areas of growth. However, it's important to note that the MBTI has also received criticism for its lack of scientific evidence, limited predictive power, and potential for oversimplification of complex human personalities

VI. CONCLUSION

The Personality Prediction System using ML and NLP is a beneficial tool for various entities in their field of work. By leveraging ML algorithms and NLP techniques, the system can analyse textual data and extract valuable insights into an individual's personality traits, characteristics, and behaviours.

ML models trained on the dataset of textual and mcq responses have effectively captured the nuances and complexities of personality, allowing for accurate predictions and understanding of individual differences. NLP techniques enabled the analysis of semantic content, contextual information, and even subconscious cues present in text, providing a comprehensive view of personality dynamics.

Overall, Personality Prediction through ML and NLP represents a powerful approach for unravelling the intricacies of human personality. As technology advances and research in this area progresses, we can anticipate further advancements in our understanding of personality traits, leading to more accurate predictions and valuable insights that can benefit diverse fields and contribute to our knowledge of human behaviour.

VI. REFERENCES

[1] R. Hegde, S. Hegde, Sanjana, S. Kotian, and S. Shetty, “Personality classification using Data mining approach,” IJRAR, 2019.

[2] F. Q. Annisa, E. Supriyanto and S. Taheri, "Personality Dimensions Classification with EEG Analysis using Support Vector Machine," 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 2020, pp. 79-82, doi: 10.1109/ISRITI51436.2020.9315507.

[3] D. Al-Hammadi and R. K. Moore, "Using Sampling Techniques and Machine Learning Algorithms to Improve Big Five Personality Traits Recognition from Non-verbal Cues," 2021 National Computing Colleges Conference (NCCC), Taif, Saudi Arabia, 2021, pp. 1-6, doi: 10.1109/NCCC49330.2021.9428804.

[4] Alam Sher Khan, Hussain Ahmad, Muhammad Zubair Asghar, Furqan Khan Saddozai, Areeba Arif and Hassan Ali Khalid, “Personality Classification from Online Text using Machine Learning Approach” International Journal of Advanced Computer Science and Applications (IJACSA), 11(3), 2020.

[5] Z. Mushtaq, S. Ashraf and N. Sabahat, "Predicting MBTI Personality type with K-means Clustering and Gradient Boosting," 2020 IEEE 23rd

International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 2020, pp. 1-5, doi: 10.1109/INMIC50486.2020.9318078

[6] Sagar Patel, Mansi Nimje, Akshay Shetty, Sagar Kulkarni, 2021, "Personality Analysis using social media", INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NTASU – 2020 (Volume 09 – Issue 03)

[7] M. N. Sahono et al., "Extrovert and Introvert Classification based on Myers-Briggs Type Indicator (MBTI) using Support Vector Machine (SVM)," 2020 International Seminar on Application for Technology of Information and Communication (semantic), Semarang, Indonesia, 2020, pp. 572-577, doi: 10.1109/iSemantic50169.2020.9234288.

[8] V. dos Santos, I. Paraboni, "Myers-Briggs personality classification from social media text using pre-trained language models". JUCS - Journal of Universal Computer Science, 28, 4. 378–395, 2022. doi: 10.3897/jucs.70941.

[9] E. J. Zaferani, M. Teshnehlab and M. Vali, "Automatic Personality Perception Using Autoencoder and Hierarchical Fuzzy Classification," 2021 26th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran, 2021, pp. 1-7, doi: 10.1109/CSICC52343.2021.9420627.