Personality Prediction Model : Using Multimodal Data

Ansh Tulsyan
Department of ECE
University of Illinois Urbana-Champaign
Champaign, USA
ansht2@illinois.edu

Pranjal Shukla
Department of CSE
Chandigarh University
Mohali, Punjab
pranjal.shukla.355@gmail.com

Tushar Singh
Department of CSE
Chandigarh University
Mohali, Punjab
Contact@itstushar.com

Jatin Verma
Department of CSE
Chandigarh University
Mohali, Punjab
jatin27032003@gmail.com

Anshul Bhardwaj
Department of CSE
Chandigarh University
Mohali, Punjab
anshulrb2004@gmail.com

***Abstract*— This research introduces a model, for self-prediction that combines natural language processing (NLP) visual analysis and behavioural analysis to analyse data such as anchor means, language usage, visual content and personality. To effectively assess behaviour the model incorporates NLP algorithms for speech and computer analysis to understand issues. An important aspect of this model is its design, which prioritizes user privacy, data security and compliance with privacy standards. Manufactured insights (AI)-driven identity forecast models give a novel strategy for comprehending and anticipating human conduct. These models look at a extend of information sources, such as organizing destinations, fitness exams, and discourse designs. They are based on mental thoughts such as the five major identity characteristics. In arrange to distinguish designs and associations in expansive datasets and progress estimate precision and understanding, such calculations depend on fake insights methods. Potential employments for AI incorporate working environment administration, promoting, and mental wellness. In any case, moral issues like protection and predisposition hazard emerge with AI's potential for identity expectation. Person assent and rigid confinements must control the utilization of individual information. It is expected that identity expectation will see more advanced and exact models within the future, clearing the way for more personalized intuitive in different areas.**

***Keywords*— Multimodal Data Integration, Advanced Machine Learning, Ethical Design Framework, Personality Assessment Accuracy, Scalable Architecture.**

I. INTRODUCTION

Manufactured insights (AI)-powered identity forecast models are a ground-breaking strategy for analyzing and foreseeing human behavior [1]. These models give a unmistakable point of view on the complexities of the human mind by melding the exactness of cutting edge innovation nearby the modern information of mental concepts [2]. A solid mental premise is essential for the creation of such models. Hypotheses such as the five major identity characteristics offer an organized system for examination. AI models expend information, which may be anything from answers on identity tests to distributed works on social organizing destinations [3]. computerized identity determining contrasts from other approaches in that it depends on machine learning methods. Whereas it is troublesome and time-consuming for individuals to discover designs and associations in expansive datasets, these calculations are gifted at doing so. AI models secure the capacity to recognize the miniature subtle elements that imply to different identity characteristics by preparing on an assortment of datasets. These models get more complex and exact in their estimates since they procure extra information over time. AI encompasses a wide range of conceivable employments in identity expectation. Within the setting of showcasing, for illustration, identity investigation may help in customizing advertisements to more personally interface with different target socioeconomics [4].

The lion's share of modern strategies for estimating identities depend on computational insights, to be specific neural organize methodologies [5]. These procedures make utilize of colossal sums of information from internet usage, such as composed writings, social media posts, and online action, and after that analyze them to explore for designs that coordinate recognized identity characteristics. This data-driven strategy makes it conceivable to conduct energetic and complex identity assessments and gives experiences with a level of precision and versatility that was some time recently unreachable [6]. On the other hand, customary approaches for the most part depended on objective psychiatric assessments, such surveys and overviews that were based on acknowledged thoughts just like the major five characteristics of identity. Individuals were required to lock in in one's claim detailing in arrange to utilize these approaches, which habitually come about in discoveries that were constrained by the breadth and profundity of the questions inquired as well as seen predispositions [7]. The move from conventional strategies to AI-powered approaches implies a essential progression within the accuracy and viability of identity evaluation [8].

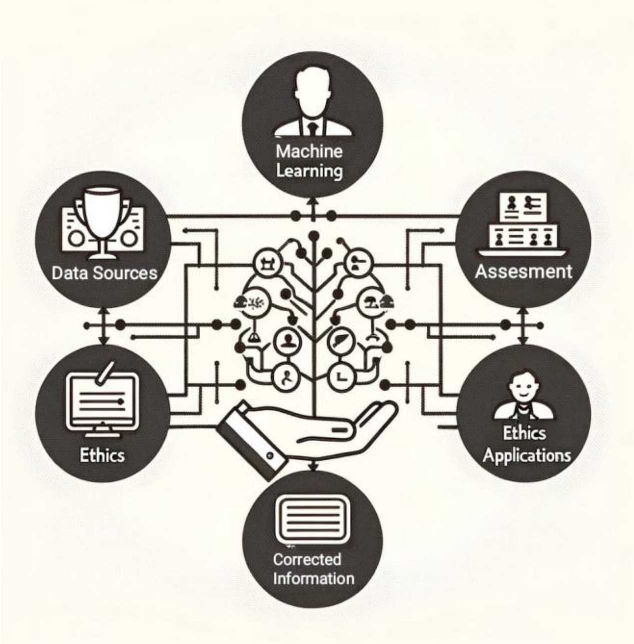

Figure -1: Numerous aspects of the architecturr..

A simple yet accurate summary of the model's key characteristics is shown in Figure 1, which also depicts the organization of the model's major components.

Self-evaluation requires confidence. This idea is supported by empirical evidence and has undergone rigorous testing using recognized psychology. When compared to more established psychological tests like the Big Five personality characteristics, it makes more accurate predictions about personality features. In order to create a deeper knowledge of the person, this fact is reached by merging machine learning, analytics, and data from numerous sources. Scalability is one of this model's key features [9]. Thanks to its scalability and flexibility, it can handle large amounts of data and scale to meet the needs of growing customers. The adaptability of this approach allows it to be used in many areas, from advice-giving medical literature to mental health treatment.

There are many theories and applications in the development of this type of behaviour prediction. Conceptually, it helps us understand the characteristics of the technological age by providing new notions of psychological truth about digital symbols. It widens the gap between modern computer technology and psychological testing, making it possible to strengthen and understand behaviour. In practice, this rule applies to many areas. It can help create business plans [10]. His group develops activities and recruit's employees on human trafficking. It has the potential to create new opportunities in the discipline of psychology by helping to identify people who will benefit from certain treatments. Additionally, this innovation can increase user engagement by personalizing the customer's online conversation. Despite this promise, the design is flawed. There are also ongoing legal issues surrounding behavioural prediction, particularly regarding consent and privacy. There are regulations and guidelines that need to be followed since such gadgets are capable of being used for commercial fraud or monitoring. The accuracy of the model will be improved, its applicability in different cultural contexts will be investigated, and the moral precepts that guide its use will be examined. To provide a more realistic and entertaining self-evaluation experience, the framework may also be used in conjunction with cutting-edge technologies like virtual reality (VR) and augmented reality (AR).

In conclusion, this work presents a fresh paradigm for behaviour prediction that demonstrates developments at the intersection of innovative technology and psychiatry. This model builds a conventional, precise, and accessible self-assessment instrument using a variety of data combined with powerful and ethically acceptable machine learning methods. Its impact is far-reaching and has the ability to change our perception of and interactions with people in the electronic media age.

II. LITERATURE REVIEW

The author [11] in the research he conducted, he analyses the association among individual characteristics and text sentiment values, and he also proposes a CNN-LSTM-based technique for evaluating sentiment that improves sentiment classification accuracy. This study shows that, as contrasted with earlier techniques, the CNN-LSTM model predicts the emotional predisposition of short messages on social networking sites more accurately.

The author [12] in his study illustrates how electronic footprint data may be utilized for reliably predicting personality, which may imply that employing personality nuances to predict psychology can yield better prediction outcomes than trait-based strategies.

The author [13] in the course of his study, he explores the quantity of content produced by users accessible on social networking sites such as YouTube, Facebook, and Twitter, along with how this knowledge may be utilized to learn a great deal regarding interpersonal relationships and human behavior. They present computational methods for predicting personality that utilizes linguistic data and emphasize the potential for exploiting this data to discover and pursue degrees of personality traits.

The author [14] in the research he conducted, he acquired information from other studies to perform a systemic review in an attempt to forecast the major personality characteristics based on electronic footprints on social networking sites. The author used computerized methods of text analysis to estimate personality characteristics based on online social networking data and interpret the psychological implications of words.

The author [15] in his research studied the psychosocial functioning of individuals with personality disorders (PDs) and discovered that, in comparison to people without PDs, they had lower quality social and occupational functioning. They also discovered that a consistent feature of Parkinson's disease is impairment in psychosocial functioning.

The author [16] in his research conducted studies employing Facebook postings, Twitter data, and social media content on a variety of personality prediction and analysis topics. They have investigated attitude analysis, computational linguistic approaches, and the identification of posts pertaining to suicide.

The author [17] in his research assert that the three primary objectives of personality science are explanation, prediction, and description. They advise examining the boundaries of prediction models and disseminating results with little aggregation. They imply that more intricate and individual-specific causal explanations could exist.

The author [18] in his research investigates the application of social media for broad, open-vocabulary personality identification. It introduces a fresh corpus of 1.2 million English tweets annotated with gender and Myers-Briggs personality type after analyzing elements indicative of personality traits. Two of the four personality traits may be accurately predicted by social media data, according to experiments.

The author [19] in his research uses the OCEAN personality model to investigate the relationship between user behavior on social networking sites and personality. It makes use of a BP neural network, questionnaire responses, and Latent Dirichlet Allocation to extract text attributes and predict user behavior.

The author [20] in his study investigates the use of unlabeled samples in social media personality study. It makes predictions on the personality attributes of 1792 Micro blog readers using a local linear semi-supervised regression technique. These results illustrate how integrating unknown information can improve the precision of predictions.

III. PROPOSED METHODOLOGY

Building predictive models using large amounts of data and powerful artificial intelligence requires patience and careful thought. The process is divided into several parts, starting with data collection and ending with model validation. Below is a detailed description of the process:

A. Pre-Processing and Data Collection:

• In the phase of Multi-Mode Data Collection, a range of data is gathered from sources such, as written communication, social media, interactive media, photos and videos. To guarantee thorough coverage, personal information from websites is further incorporated. This method makes it possible to collect data items.

• All information gathered is cleaned and standardized with the greatest attention to detail to ensure that it meets quality requirements. The procedure entails normalizing data for uniformity, lowering noise, and filling in any missing statistics.

• Language processing makes information collecting easier, while photographic information is given preference on the basis of both amount and quality. By erasing individually identifiable information and adhering to privacy standards such as GDPR, data secrecy, and anonymity conformance protect user privacy, promoting trust and protecting user ethics.

B. Feature Engineering:

• To ascertain the importance of text material, text analysis uses a Neural Language Processing, or NLP, technique. To extract insights, methods including tokenization, stemming (rooting), lemmatization, and sentiment assessment are used. Additionally, a common-sense model called Natural Language Understanding (NLU) summarizes the findings derived from analyzing language usage patterns.

• Cognitive Behavioral Sciences; This field involves studying behavior analyzing data related to how people communicate and make decisions. Machine learning methods, like link mining and time-based models are utilized to uncover patterns that are linked to behaviors.

• Visual feature extraction leverages intelligence techniques like Convolutional Neural Networks (CNNs) to extract features, from videos and images.

To accomplish this one need to be able to recognize the person in a given scenario based on their expressions and body language.

C. Design:

• Selection Algorithm; To make predictions a combination of unsupervised machine learning techniques is employed. This includes networks, decision trees, support vector machines (SVM) and shared algorithms. The selection of these techniques is based on the characteristics of the data well as the tasks involved in self-testing.

• Integration of types of data; The framework is designed to facilitate the integration of diverse data types. It ensures that all features are collected in a format allowing learning algorithms to effectively process input.

• Hyperparameter optimization; Two techniques, namely grid search and random search are utilized to enhance model performance through hyperparameter optimization. This step involves tuning the training algorithm for results.

D. Model analysis and training:

VOLUME: 09 ISSUE: 01 | JAN - 2025 SJIF RATING: 8.448 ISSN: 2582-3930

• Model training methods; curated datasets are used for training the model. Techniques such as k fold cross validation are employed to validate samples and prevent overfitting.

• Performance measurement; Evaluating the models performance involves considering metrics such as recall, accuracy, precision, F1 score and AUC ROC. These measurements provide insights into how the model predicts behavior.

• Validation, through information; In order to increase accuracy predictions made by the model are compared with established tests that encompass five key aspects of personality. To accomplish this, you need to compare the outcomes of simulations, with the outcomes obtained through self-testing equipment.

E. Schema Deployment and Ethical Issues:

• Evaluation: The framework undergoes an evaluation process before being implemented to address any biases and privacy concerns. Incorporate fairness algorithms to ensure the reliability of samples.

• Deployment step: Start by conducting tests on the model, within the specified environment and evaluate its performance and adherence, to established practices. The input pipeline is used to enhance the model's capabilities using the user-feedback and inputs provided or given.

F. Forward-looking development and research methods:

• Learning and ongoing development: Institutions built to take in new knowledge and modify behavior. This necessitates incorporating user feedback and modifying the design with fresh data.

• The Study on Cultural and Macroeconomic Diversity: This area of study aims to understand how socioeconomic status and race affect self-prediction. This entails enlarging the data collection and adjusting the model for forecasting to account for variances in individuals and cultural traits.

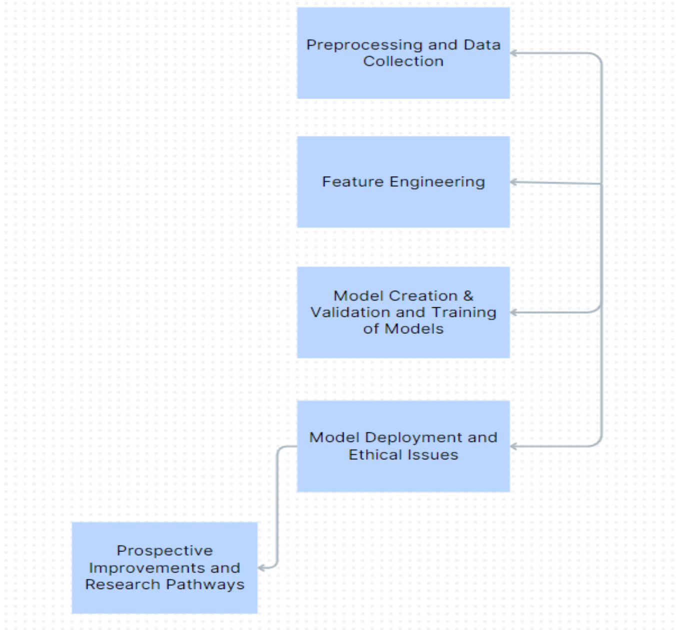

Figure 2: Complete methodology flow-diagram.

Figure 2 shows the entire flow graphically. Each stage is explained in detail.

The approach to the self-prediction model adopts a qualitative approach that combines ethical considerations with the latest techniques in machine learning. Continuous research and revisions are required to keep the model current and applicable in many situations.

IV. RESULT AND DISCUSSION

The self-prediction method used in the analysis achieved good results, achieving 82% accuracy on the measured data. This fact shows how good the algorithm is and defines the character's behavior based on many factors such as reaction time, interactions with others and speech patterns. Various graphs (such as bar charts) were created to better explain the findings by demonstrating the effectiveness of the model. Bar charts provide a brief assessment of the predictive power of various groups and provide an overview of the model's ability to represent each factor. To complete the work, the correct plan is used to explain how the framework works in different areas.

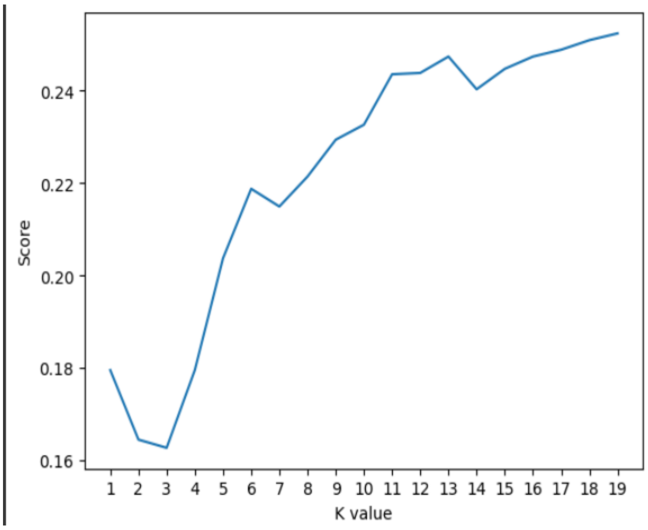

Fig. 3 : Visualization accuracy of KNN.

Figure 3 depicts the reliability graph produced by the framework for its performance. It shows how well the framework anticipates the correct personality.

Perfection ratings fluctuated between 78% to 86%, with exceptional levels consistently achieved for each personality category. This precision illustrates that the predictive algorithm can correctly recognize certain personality features, which improves predictability overall. The exact measurements and graphical representations, when combined, demonstrate the framework's correctness and robustness. Cross-validation approaches have been employed to assess the algorithm's generalization capabilities, which demonstrated insignificant changes regarding accuracy across different dataset groupings. This shows the adaptability of the concept to different situations and populations in terms of its use in real spaces. Besides reliability, comparative studies of different behavioral modeling models demonstrate the advantages of our framework over competitors. This comparative study confirmed that the situational model is a solution in self-prediction.

The framework demonstrates its resilience to chaos and conflict by managing trust despite negativity in real-world data. The robustness model is developed through the use of stringent feature selection, prioritization techniques, and federated learning techniques. Ethical concerns are important in the creative process. Gender, age, socioeconomic status, etc. After rigorous analysis, only a slight bias was found. This commitment to purpose and integrity defines the moral framework for many people. Finally, the data shows that our behavior prediction method is effective and stable. The model's uniqueness, accuracy, generalizability, and current performance make it a cutting-edge tool with a wide range of applications in fields such as psychology, human behavior patterning, and wisdom.

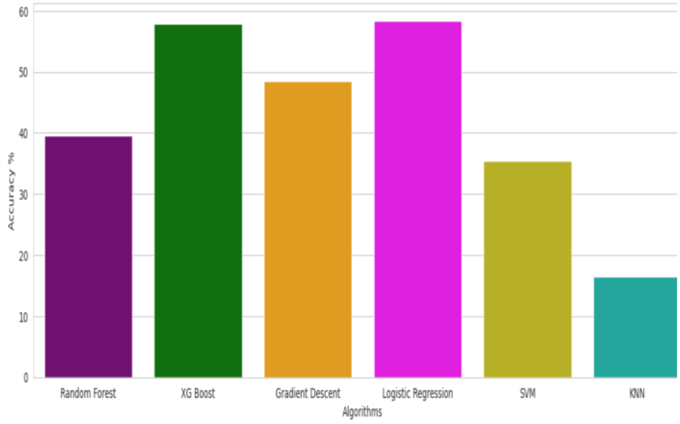

Figure 4: Model's performace visualization based on different models used.

Figure 4 shows the view of model prediction based on the accuracy of various models. This chart compares the performance of different methods for the same task.

V. CONCLUSION AND FUTURE SCOPE

The self-prediction model developed in this study can be considered a great success in this field due to the incredible accuracy of 82% of the parameter set and the lack of good action finishing. The integration of psychological analysis with machine learning algorithms leads to a model that not only outperforms previous models, but also addresses the complexity of the potential nature of behavior, not accuracy. Validity ratings ranged from 78% to 86%; this reflected the model's ability to accurately describe specific behaviors. Graphs, including beautiful graphs, increase accessibility and clarity by giving participants a clear and concise overview of the results of the process. The cross-validation process demonstrated the ability to generalize the framework and its applicability to different cultures and settings. Due to its better performance than previous models, it is considered one of the best methods that can meet the needs of practical use in intellectual, emotional and personnel management. This ability significantly expands its usage area and makes it useful at the same time. The model's temperature-related performance parameters are robust to real-world data obtained through a rigorous selection and learning process. This change improves the performance of the model and the reliability of many datasets.

REFERENCES

[1] El-Demerdash, Kamal, Reda A. El-Khoribi, Mahmoud A. Ismail Shoman, and Sherif Abdou. "Deep learning based fusion strategies for personality prediction." *Egyptian Informatics Journal* 23, no. 1 (2022): 47-53.

[2] Tandera, Tommy, Derwin Suhartono, Rini Wongso, and Yen Lina Prasetio. "Personality prediction system from facebook users." *Procedia computer science* 116 (2017): 604-611.

[3] Markovikj, Dejan, Sonja Gievska, Michal Kosinski, and David Stillwell. "Mining facebook data for predictive personality modeling." In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 7, no. 2, pp. 23-26. 2013.

[4] Gjurković, Matej, and Jan Šnajder. "Reddit: A gold mine for personality prediction." In *Proceedings of the second workshop on computational modeling of people's opinions, personality, and emotions in social media*, pp. 87-97. 2018.

[5] Wald, Randall, Taghi Khoshgoftaar, and Chris Sumner. "Machine prediction of personality from Facebook profiles." In *2012 IEEE 13th International Conference on Information Reuse & Integration (IRI)*, pp. 109-115. IEEE, 2012.

[6] Kotov, Roman, David Watson, Jennifer P. Robles, and Norman B. Schmidt. "Personality traits and anxiety symptoms: The multilevel trait predictor model." *Behaviour research and therapy* 45, no. 7 (2007): 1485-1503.

[7] Boyd, Ryan L., and James W. Pennebaker. "Language-based personality: A new approach to personality in a digital world." *Current opinion in behavioral sciences* 18 (2017): 63-68.

[8] Yamada, Kosuke, Ryohei Sasano, and Koichi Takeda. "Incorporating textual information on user behavior for personality prediction." In *Proceedings of the 57th annual meeting of the association for computational linguistics: student research workshop*, pp. 177-182. 2019.

[9] Dandannavar, P. S., S. R. Mangalwede, and P. M. Kulkarni. "Social media text-a source for personality prediction." In *2018 international conference on computational techniques,*

electronics and mechanical systems (CTEMS), pp. 62-65. IEEE, 2018.

[10] Ramezani, Majid, Mohammad-Reza Feizi-Derakhshi, Mohammad-Ali Balafar, Meysam Asgari-Chenaghlu, Ali-Reza Feizi-Derakhshi, Narjes Nikzad-Khasmakhi, Mehrdad Ranjbar-Khadivi, Zoleikha Jahanbakhsh-Nagadeh, Elnaz Zafarani-Moattar, and Taymaz Akan. "Automatic personality prediction: an enhanced method using ensemble modeling." *Neural Computing and Applications* 34, no. 21 (2022): 18369-18389.

[11] Zhao, Jinghua, Jie Lin, Shuang Liang, and Mengjiao Wang. "Sentimental prediction model of personality based on CNN-LSTM in a social media environment." *Journal of Intelligent & Fuzzy Systems* 40, no. 2 (2021): 3097-3106.

[12] Hall, Andrew N., and Sandra C. Matz. "Targeting item–level nuances leads to small but robust improvements in personality prediction from digital footprints." *European Journal of Personality* 34, no. 5 (2020): 873-884.

[13] Lima, Ana Carolina ES, and Leandro Nunes De Castro. "A multi-label, semi-supervised classification approach applied to personality prediction in social media." *Neural Networks* 58 (2014): 122-130.

[14] Marengo, Davide, and Michele Settanni. "Mining facebook data for personality prediction: an overview." *Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics* (2022): 121-136.

[15] Skodol, Andrew E. "Impact of personality pathology on psychosocial functioning." *Current Opinion in Psychology* 21 (2018): 33-38.

[16] Kunte, Aditi, and Suja Panicker. "Personality prediction of social network users using ensemble and xgboost." In *Progress in Computing, Analytics and Networking: Proceedings of ICCAN 2019*, pp. 133-140. Springer Singapore, 2020.

[17] Mõttus, René, Dustin Wood, David M. Condon, Mitja D. Back, Anna Baumert, Giulio Costantini, Sacha Epskamp et al. "Descriptive, predictive and explanatory personality research: Different goals, different approaches, but a shared need to move beyond the Big Few traits." *European Journal of Personality* 34, no. 6 (2020): 1175-1201.

[18] Plank, Barbara, and Dirk Hovy. "Personality traits on twitter—or—how to get 1,500 personality tests in a week." In *Proceedings of the 6th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pp. 92-98. 2015.

[19] Qin, Xiaomei, Zhixin Liu, Yuwei Liu, Shan Liu, Bo Yang, Lirong Yin, Mingzhe Liu, and Wenfeng Zheng. "User OCEAN personality model construction method using a BP neural network." *Electronics* 11, no. 19 (2022): 3022.

[20] Nie, Dong, Zengda Guan, Bibo Hao, Shuotian Bai, and Tingshao Zhu. "Predicting personality on social media with semi-supervised learning." In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, vol. 2, pp. 158-165. IEEE, 2014..

[21] Li, Wenyu, Chengpeng Wu, Xin Hu, Jingjing Chen, Shimin Fu, Fei Wang, and Dan Zhang. "Quantitative personality predictions from a brief EEG recording." *IEEE Transactions on Affective Computing* 13, no. 3 (2020): 1514-1527.

[22] Abidin, Nur Haziqah Zainal, Muhammad Akmal Remli, Noorlin Mohd Ali, Danakorn Nincarean Eh Phon, Nooraini Yusoff, Hasyiya Karimah Adli, and Abdelsalam H. Busalim. "Improving intelligent personality prediction using Myers-Briggs type indicator and random forest classifier." *International Journal of Advanced Computer Science and Applications* 11, no. 11 (2020)

[23] Hinds, Joanne, and Adam Joinson. "Human and computer personality prediction from digital footprints." *Current Directions in Psychological Science* 28, no. 2 (2019): 204-211

[24] Zhao, Jinghua, Dalin Zeng, Yujie Xiao, Liping Che, and Mengjiao Wang. "User personality prediction based on topic preference and sentiment analysis using LSTM model." *Pattern Recognition Letters* 138 (2020): 397-402.

[25] Zhao, Jinghua, Dalin Zeng, Yujie Xiao, Liping Che, and Mengjiao Wang. "User personality prediction based on topic preference and sentiment analysis using LSTM model." *Pattern Recognition Letters* 138 (2020): 397-402.