A Literature Review in Personality Predictions Based on Twitter Text Modality

Tharsika Thurairasa
Information Technology
SLIIT
Colombo, SriLanka
tharsikasliit@gmail.com

Lakmal Rupasinghe
Computer Systems Engineering
SLIIT
Colombo, SriLanka
lakmal.r@sliit.lk

***Abstract* — Automatic personality detection of an individual's character qualities has numerous significant useful applications. Personality assessment is used to evaluate the individual on different aspects. With regards to assumption examination, for instance, the items and administrations prescribed to an individual ought to be those that have been emphatically assessed by different users with a comparable Personality type. This paper gives an overview of the advancement of personality prediction from twitter text modality, the regular issues looked in performing said task, and further upgrades that can be applied later on.**

Keywords* — *personality prediction, social media, machine learning, deep learning, big five model

I. INTRODUCTION

In these days computerized personality detection is another and forthcoming field. There have not been numerous extensive writing reviews done in character location and our paper is the first which gives the reader a bird's-eye perspective. It gives the idea about on the ongoing patterns and advancements in the field. Texts often reflect different aspects of the author's personality [4]. There is no ongoing work which gives the reader a general point of view of the advances mechanized personality detection specially using text dataset. As indicated by the Merriam-Webster dictionary the social media is characterized as types of electronic communication. Online media is an inescapable aspect of the Internet, as insights show that individuals burn through 1 in like clockwork of their Internet use via social media. A perception with respect to Facebook use detailed that clients sign into their Facebook accounts from 2 to 5 times each day, with a normal of 5 - 15 minutes for every meeting [14].

II. PERSONALITY MEASURES

A. *Personality Measures Theories*

Personality Theories fall into fundamental 4 classifications.

Psychoanalytic Theory: - As indicated by Freud, the per sodality is comprised of three segments as id, inner self and superego. id alludes to the drive vitality that is liable for the human needs like sustenance, thankfulness and urges like scorn, desires and so on. The superego or soul, represent profound quality and Social standards, speak to what an individual need to be. Conscience the third part chips away at the guideline of reality that intervenes between the requests of the primary segment id and the second segment superego and afterward picks the most reasonable answer as long as possible [1].

Trait Theory: - Present-day attribute hypothesis attempts to show character by setting of various grouping measurements (generally following a lexical methodology) and developing a poll to measure them. Analysts have utilized different plans for personality demonstrating, for example, 16PF, EPQ-R and three quality character model PEN where there are super-factors Psychoticism, Extraversion, and Neuroticism (PEN) at the head of the pecking order. The Myers–Briggs Type Indicator (MBTI) is one of the most generally managed individuality tests, given a huge number of times each year to representatives in a great many organizations. The MBTI character measure arranges individuals into two classifications in every four dimensions: self-preoccupation versus extraversion, detecting as opposed to intuiting, thinking as opposed to feeling, and judging as opposed to seeing. The most mainstream measure utilized in the writing on mechanized character discovery is by a wide margin the Big-Five-character qualities, which are the accompanying paired (yes/no) values [2].

Humanistic Theory: - Maslow accepted that character depends on close to home decisions not on nature or sustain. He recommended that individuals have and are inspired to assist them with pursuing their necessities or want that was spoken to in and the last level: self-realization that is creating and developing to arrive at genuine potential [1].
Social Cognition Theory: - The social insight hypothesis see character in the type of social communications. An individual's conduct is influenced by nature in which he remains [1].

The Trait hypothesis is most broadly utilized in examining the character in the field of Psychology. Dissimilar to different speculations this depends on finding the contrasts between the characters of people. The blend of different characteristics frames a character that is consistently one of a kind for each person. Most examinations on character forecast have zeroed in on the Big Five or MBTI personality models, which are the two most utilized personality models.

Big Five: The present Researchers accept that there are 5-character qualities. Large Five recommends that the attributes can be ordered on 5 unique classes. Careful marks for these 5 characteristics are as yet hard to concur for some of them. The famous abbreviation is OCEAN for traits.

Extraversion (EXT): - Individuals with high Extroversion quality have high certainty, Positive vitality, and positive feelings, friendly and inclination to communicate more with others. They are garrulous in nature. It repudiates saved conduct. Components identified with this quality are vitality,

ISSN 2756-9160 / November 2020.
International Conference on Advances in Computing and Technology (ICACT–2020) Proceedings

garrulity, carefree, well disposed, helping and so forth. These individuals like themselves just as about their general surroundings. Individuals with low extroversion are saved, calm.

Neuroticism (NEU): - It is repudiated sure or secure nature. individuals with high neuroticism touchy or apprehensive. This characteristic is described by trouble, grouchiness, and passionate insecurity. They experience negative feelings and feelings effectively, similar to outrage, uneasiness, sorrow, pessimism and so forth. It alludes to the inclination to encounter negative enthusiastic states and see oneself and the world around contrarily. Variables like volatile, on edge on edge and so on are some related attributes.

Agreeableness (AGR): - This is the propensity to be agreeable with others as opposed to being dubious. They are benevolent and enjoyed by their associates just as individuals encompassing them. They don't care to battle or contend as opposed; they are harmony creators. Humble, amenability, supportive, understanding, kind, touchy and so forth are the qualities that go under the umbrella of suitability.

Conscientiousness (CON): - It alludes to the fitness of being consistent, self-taught, capable, zeroing in on accomplishing objectives, and organizes designs rather than unconstrained conduct. It contrasts indiscreet conduct. It means how cautious, careful, genuine an individual is. It is an approach to control driving forces and act in a manner that is adequate socially by everybody around. These individuals are extraordinary at arranging and sorting out viably. These incorporate components as arranging, capable, difficult work, assurance, eager, control and so forth. They are acceptable in administration characteristics.

Openness (OPN): - It mirrors the scholarly degree of an individual. How inquisitive, innovative novel an individual is. It likewise reflects how creative or autonomous an individual is. Transparency is identified with individuals' enthusiasm to attempt to new things, the capacity to be defenseless, and the ability to consider new ideas. Basic characteristics identified with transparency are: Imagination, different interests, Originality, Daring, Cleverness, Intellect, Creativity, Curiosity and so forth.

III. Approaches used

There is a critical developing enthusiasm for computerized personality prediction utilizing web-based media among specialists in both the Natural Language Processing and Social Science fields. Up until now, the utilization of customary personality tests has generally been restricted to clinical brain research, advising and human asset the board. Be that as it may, computerized character forecast from web-based media has a more extensive application, for example, online media advertising or dating applications and sites [3].

There have been numerous techniques utilized for personality detection as demonstrated as follows.

Questionnaire: - The most punctual type of approach utilized for character expectation was in the type of inquiries. Clients were posed a few inquiries that had various options, from which the client needed to choose one. These Questions were distinctive for various character qualities.

Semantic Similarity: - In this for the characteristics there are pre-characterized jargon or word reference words. The client's words present in the posts are checked for the semantic comparability, for example, comparative implications have the same score. The separation is discovered and hence the attribute was anticipated.

Machine Learning: - Classical approaches can't deal with the vast amount of data. This is one of the upsides of Machine learning calculations. Machine learning algorithms utilize computational techniques to "learn" data legitimately from information without depending on a foreordained condition as a model.

Deep Learning: - Models are prepared by utilizing an enormous arrangement of marked information and neural network architectures that contain numerous layers. Deep learning can be utilized to foresee the character attributes with more precision. It measures a similar path as human minds do. The component extraction measure is and there is no over-burden [1].

Table 4 : Popular Datasets, Model, Methodology

Dataset and Model	Methodology
Big Five, MyPersonality	Deep learning architecture - AttRCNN
Big Five, MyPersonality	Neural Networks like CNN, RNN
Big Five, MyPersonality	XGBoost, Logistic Regression, SVM
Big Five, MyPersonality	LIWC, Semi supervised, PMC

IV. Applications

Employment screening: - In human assets the executives, character attributes influence one's reasonableness for specific positions.

Criminology: - If the police know about the character qualities of the individuals who were available at the wrongdoing scene, it might help in lessening the hover of suspects.

Specific medical care and guiding: - As of 2016, almost 33% of Americans have looked for proficient directing for psychological well-being connected issues.

Word extremity location: - Personality recognition can be misused for word extremity dis-ambiguation in estimation vocabularies, as a similar idea can pass on various significance to various kinds of individuals.

Suggestion frameworks: - People that share a specific character type may have similar interests and diversions.

Upgraded Personal Assistants: - Present-day robotized voice aides, for example, Siri, Google Assistant, Alexa.

V. Personality Prediction from text review

Personality detection is where data about a person's character attribute is recognized, given a lot of information. There have been a few ways to deal with automated character

ISSN 2756-9160 / November 2020.
International Conference on Advances in Computing and Technology (ICACT–2020) Proceedings

prediction dependent on various types of datasets, for example, social media post, face Tube, speech, smartphone, video, essays, handwriting, travel pattern, gender, age. This paper will mainly focus to review personality prediction using text dataset.

A. Baseline Methods for Text

The following subsections summarize the models, dataset and techniques which had been used in machine learning, deep learning-based personality detection on text modality.

a) Twitter - text Dataset

Most of the personality prediction research studies was attempted on Twitter. In the year of 2016, the authors in this paper [5] This study had used text classification to predict personality based on text written by Twitter users. Dataset consists of last 1, 000 texts in the form of tweets and re-tweets. Collection of tweets from users is also made into a single document/ one long string, then it was preprocessed and labeled according to Big Five personality dimensions. The languages used for this study are English and Indonesian. Classification methods had implemented. Those are Naive Bayes, K-Nearest Neighbors and Support Vector Machine. Finally testing conducted using 10-fold cross-validations. Testing results showed Naive Bayes method was slightly outperformed the other methods.

In the year of 2017, the authors in this paper [6] The aim of this research is to analyze how twitter (dataset) can be utilized to improve the user experience in character assessment. propose a manner by which the client's character can be anticipated through information mapping accessible to general society on their own twitter utilizing DISC (Dominance, Influence, Compliance, Steadiness) assessment. Text mining and sentiment analysis had performed for every user dependent on his/her ongoing tweets. Downloaded more than 1,000,000 tweets utilizing catchphrases.

In the year of 2017, the authors in this paper [7] Twitter user's data for Bahasa Indonesia, the native language of Indonesia. It's possible without a tool with predefined words (LIWC, MRC) but by assessing the user's choice of words. The personality prediction framework based on Support Vector Machine and XGBoost prepared with 329 instances of users. Assessment results utilizing 10-fold cross-validation shows that the framework figured out how to arrive at the most elevated normal exactness with SVM and XGBoost. To build personality prediction used the five-factor model which also known as big five model. This framework built on XGBoost managed to perform significantly better than on SVM.

In the year of 2018, the authors in this paper [8] presented optimization techniques for automatic personality recognition based on Twitter in Bahasa Indonesia. Evaluated a progression of techniques implementing hyperparameter tuning, feature selection, and sampling to improve the machine learning calculations utilized. The personality forecast framework is based on machine learning algorithms and used big five model. There are three machine learning calculations utilized in this study, to be specific Stochastic Gradient Descent (SGD), and two ensemble learning calculations, Gradient Boosting (XGBoost), and stacking (super learner). By executing this arrangement of optimization strategies, the current examination's assessment results show immense improvement by accomplishing 1.0 ROC AUC score with SGD and Super Learner.

VI. CONCLUSION

This paper gave an understanding on existing endeavors of the task of personality prediction from twitter text dataset to-date, alongside the different sorts of twitter dataset which have been used for said task. A portion of these strategies utilize a closed-vocabulary approach with psycholinguistic devices. Recently have taken a stab at applying semi- supervised and unsupervised learning out how to handle this issue. Further enhancements to the current condition of personality prediction can be made by extending the objective language, applying more appropriate algorithms or preprocessing strategies to accomplish higher accuracy. In future robust deep learning models can build then accuracy can be improved.

REFERENCES

[1] Hetal Vora, Mamta Bhamare, and Dr. K. Ashok Kumar, "Personality Prediction from Social Media Text: An Overview," *Int. J. Eng. Res.*, vol. V9, no. 05, pp. 352–357, 2020, doi: 10.17577/ijertv9is050203.

[2] Y. Mehta, N. Majumder, A. Gelbukh, and E. Cambria, "Recent trends in deep learning based personality detection," *Artif. Intell. Rev.*, vol. 53, no. 4, pp. 2313–2339, 2020, doi: 10.1007/s10462-019-09770-z.

[3] hernandez and knight, "Predicting MBTI from text."

[4] N. Majumder, I. P. Nacional, A. Gelbukh, and I. P. Nacional, "Deep Learning-Based Document Modeling for Personality Detection from Text," 2017.

[5] B. Y. Pratama and R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM," *Proc. 2015 Int. Conf. Data Softw. Eng. ICODSE 2015*, pp. 170–174, 2016, doi: 10.1109/ICODSE.2015.7436992.

[6] N. Ahmad and J. Siddique, "Personality Assessment using Twitter Tweets," *Procedia Comput. Sci.*, vol. 112, pp. 1964–1973, 2017, doi: 10.1016/j.procs.2017.08.067.

[7] V. Ong *et al.*, "Personality prediction based on Twitter information in Bahasa Indonesia," *Proc. 2017 Fed. Conf. Comput. Sci. Inf. Syst. FedCSIS 2017*, vol. 11, pp. 367–372, 2017, doi: 10.15439/2017F359.

[8] G. Y. N. N. Adi, M. H. Tandio, V. Ong, and D. Suhartono, "Optimization for Automatic Personality Recognition on Twitter in Bahasa Indonesia," *Procedia Comput. Sci.*, vol. 135, pp. 473–480, 2018, doi: 10.1016/j.procs.2018.08.199.

[9] J. Yu and K. Markov, "Deep learning based personality recognition from Facebook status updates," *Proc. - 2017 IEEE 8th Int. Conf. Aware. Sci. Technol. iCAST 2017*, vol. 2018-Janua, no. January, pp. 383–387, 2017, doi: 10.1109/ICAwST.2017.8256484.

[10] T. Tandera, Hendro, D. Suhartono, R. Wongso, and Y. L. Prasetio, "Personality Prediction System from Facebook Users," *Procedia Comput. Sci.*, vol. 116, pp. 604–611, 2017, doi: 10.1016/j.procs.2017.10.016.

[11] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Personality Predictions Based on User Behavior on the Facebook Social Media Platform," IEEE Access, vol. 6, pp. 61959–61969, 2018, doi: 10.1109/ACCESS.2018.2876502.

[12] Hetal Vora, Mamta Bhamare, and Dr. K. Ashok Kumar, "Personality Prediction from Social Media Text: An Overview," *Int. J. Eng. Res.*, vol. V9, no. 05, pp. 352–357, 2020, doi: 10.17577/ijertv9is050203.

[13] V. Ong, A. D. S. Rahmanto, Williem, and D. Suhartono, "Exploring personality prediction from text on social media: A literature review," *Internetworking Indones. J.*, vol. 9, no. 1, pp. 65–70, 2017.