Personality Classification Model of Social Network Profiles based on their Activities and Contents

Mervat Ragab Bakry[1], Mona Mohamed Nasr[2], Fahad Kamal Alsheref[3]

Information Systems Department, Beni-Suef University, Egypt[1, 3]

Information Systems Department, Helwan University, Egypt[2]

Abstract—Social networks have become an important part of everyday life, especially after the latest technologies such as smartphones, tablets, and laptops have become widespread. Individuals spend a lot of time on social media and express their feelings and opinions through statuses, comments, and updates which could be a way to understand and classify their personalities. The personalities in psychological science are divided into five classes according to the Big-five model (Openness, Extraversion, Consciousness, Agreeableness, and Neurotic). This model shows the key features with their weights for each personality. In this paper, a proposed model is developed for detecting the personality features from users' activities in social networks. In this model, machine learning techniques are used for predicting the personalities with a score for each Big-five model type and sorting them in descending order. The personality classification model will be useful in developing a better understanding of the user profile and specifically targeting users with appropriate advertising. Any social media network user's personality can be predicted by using their posts and status updates to get better accuracy. The experimental results of the model in this study provide an enhancement because it can predict the precise score of one user in each factor of the Big-five. The proposed model was tested on a dataset extracted from Facebook and manually classified by experts, and it achieved 89.37% accuracy.

Keywords—Psychological personality; machine learning techniques; big-five; LinearSVC

I. INTRODUCTION

In recent years, social media has surpassed email as the most extensively utilized method for communication and engagement between people [1]. As people increasingly choose to connect informally through smartphones, face-to-face engagement is becoming less common. As a result, determining a person's personality is challenging. However, because individuals spend a lot of time on social media and express their feelings and opinions through statuses, comments, and updates, what is published on social networks can help to receive the information needed for this study.

Users on Facebook typically express their views and ideas through status updates or comments. Despite the fact that Facebook is now more extensively used to post photographs and videos, this study focuses on the linguistic element of Facebook users' status updates. There is a link between a person's personality and his or her language conduct, according to studies in the field of psychology [2] and [3]. Using a natural language processing technique, this association may be successfully evaluated and shown. As a result, the purpose of this study is to develop a system that can automatically estimate a user's personality based on their Facebook behaviors.

Big Five Personality, MBTI (Myers-Briggs Type Indicator), and DISC are some of the personality models used to predict personality (Dominance Influence Steadiness Conscientiousness). However, after some thought and a literature search, Big-five personality was chosen for this paper since it is the most common and accurate method of determining a person's personality qualities. Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism are the traits in this paradigm.

The corpus used in this study consists of a one dataset which contains 100 Facebook profiles with their 27,182 posts classified over 38 categories. The rest of this paper is organized as follows. Section II gives background information about used concepts and techniques. Section III discusses related work. Section IV describes the proposed model for personality classification model. In Section V, the case study and its findings are discussed and the conclusion in Section VI.

II. BACKGROUND

Several methodologies, such as the Personality model and other classification algorithms, are employed in the proposed model, as discussed in the following sections.

A. Personality Models

Personality is a persistent tendency and attribute that governs how an individual's psychological behavior is similar and different (thoughts, emotions, behaviors). In other words, personality reflects patterns of human behavior, thought, and interpersonal communication, and has a significant impact on a variety of elements of life, including happiness, taste, and physical and mental health [3] and [4]. There are other psychological personality models, but the Big Five (shown in Fig. 1) is the most often utilized.

1) Big-five model: Many academics nowadays believe that there are five basic personality traits [5]. For several years, evidence of this notion has been growing, beginning with D. W. Fiske's (1949) research and later elaborated upon by other researchers such as Norman (1967), Smith (1967), Goldberg (1981), and McCrae & Costa (1983). (1987).
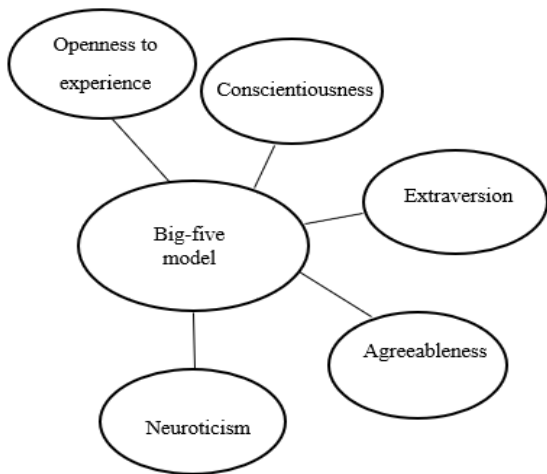
Fig. 1. Big-Five Model [5].

The big five personality traits are broad categories of personality traits are:

- Openness: Imagination and insight are two features of this attribute [5]. People with a high level of this attribute also have a diverse set of interests. They are curious in the world and other people, and they are eager to learn new skills and have a better understanding of new concerns. People with a high level of this characteristic are more likely to be daring and creative. People with low levels of this component are more traditional and may have difficulty thinking abstractly [6].

- Conscientiousness: High levels of thinking, effective impulse control, and goal-directed activities are all common aspects of this component [5]. People that are very conscientious have a propensity to be structured and detail oriented. They consider the consequences of their activities and are mindful of deadlines [6].

- Extraversion: Excitability, sociability, talkativeness, decisiveness, and extensive use of emotional emotions are all characteristics of extraversion [5]. People with a high level of extraversion are friendly and have a proclivity to gain power in social situations. Being in the company of others helps them to feel empowered and happy. People with a low level of extraversion are more likely to be timid and have less ability to consume in social contexts [6].

- Agreeableness: Characteristics like confidence, altruism, humanism, love, and other prosocial behaviours are included in this personality component [5]. People who score high in agreeableness are more helpful, whereas those who score low are more competitive and, at times, cunning [6].

- Neuroticism: Sadness, moodiness, and emotional unpredictability are all symptoms of neuroticism [5]. People with high levels of this component are more likely to experience mood swings, concern, irritability, and disappointment. People who score low on this characteristic are more emotionally balanced [6].

B. Classification Algorithms

As mentioned in section I, there are 38 categories and the collected posts and comments are classified on them, so in this study the several classification algorithms are used to classify new entries into the predefined categories or classes.

The selected classification algorithms are:

- Linear Support Vector Classification (LSVC)

- Logistic Regression (LR)

- The multinomial Naive Bayes classifier (MNB)

- Random Forest Classifier (RFC)

The accuracy is calculated for each algorithm to find the highest accuracy algorithm; the results exist in Section V.

III. RELATED WORK

Many researchers have advocated extensive customization of user personality traits and other social traits in recommender systems, and in this section, a literature survey on personality traits is presented. Personality traits are used in many areas to improve recommender systems, and many researchers have advocated extensive customization of user personality traits and other social traits in recommender systems.

Gozde et al [7], collect data from 99,217 participants from 41 countries as part of the COVIDiSTRESS worldwide survey in order to better understand the numerous relationships between COVID-19 psychological outcomes and the "big-five" personality traits. Multigroup confirmatory factor analysis and a multilevel regression model were used to examine the data. The findings revealed that throughout the pandemic, the big-five personality characteristics were strongly linked to feelings of stress and loneliness, while some of the links were weak. Their research aids in the identification of susceptible individuals and the optimization of psychological therapy during and after the COVID-19 pandemic, where neuroticism is a role in stress and loneliness vulnerabilities, particularly during disasters.

Jose et al in [8] try give a complete knowledge of the primary user privacy problems that influence DDI. The methodology used in this study is divided into three phases: (i) a systematic literature review (SLR); (ii) in-depth interviews on user privacy concerns framed through the lenses of UGD and DDI; and (iii) topic modelling using a Latent Dirichlet allocation (LDA) model to extract insights related to the object of study. They identify 14 areas linked to the research of DDI and UGD techniques based on the findings. In addition, 14 future research topics and 7 research propositions for the study of UGD, DDI, and user privacy in digital marketplaces are offered and it finished with a discussion of the importance of user privacy in DDI in digital markets. And therefore, raised the importance of understanding the personality of users, forecasting user behaviour, and evaluating their actions.

Kamal et al in [9] offered a new deep learning-based personality prediction and classification model that combines data and classifier-level fusion. Using prominent pre-trained language models such as Elmo, ULMFiT, and BERT, this approach gained from transferring learning to natural language

processing. The suggested model indicates the effectiveness of the strategy as a potential personality prediction model. Their proposed model proved to be more accurate than previous ones with 1.25 per cent and 3.12 per cent, respectively.

Valanarasu in [10] proposed a new machine learning method for predicting people's personalities based on the digital footprint of social media. During COVID-19, the suggested model may be verified for each job seeker by registering online for each organization. The suggested algorithm uses dynamic multi-context information, such as account information from Facebook, Twitter, and YouTube, among other sites. Personality prediction was more accurate than other available approaches. Even though human reasoning changes seasonally, the suggested algorithm regularly outperforms other current traditional techniques to predicting human cognition.

Masud [11] surveyed a chosen number of Facebook users based on the Short Dark Triad (SD3) model. Random Forest, Support Vector Machine, and Nave Bayes algorithms were used to categorize dark triads like Machiavellianism, Self-Centricity, and Psychopathy in this study. Compared to Random Forest and SVM, Nave Bayes produced superior results. However, showing psychopathy and self-centeredness is more difficult to detect than Machiavellianism, according to their research.

Yang Li et al in [12] proposed a new multi-task learning framework that predicted personality traits and emotional behaviours at the same time, based on the well-known correlation between personality traits and emotional attitude. It also empirically evaluated and described various information exchange mechanisms between the two tasks.

Fatemeh et al in [13] proposed a deep learning-based method for the detecting personalities from text. Convolutional neural networks (CNNs) were utilized, which have previously been found to be effective in natural language processing and personality recognition. A series of tests were used to validate their suggested technique using an essay dataset, and the empirical findings indicated that it outperformed machine learning and deep learning methods for personality detection tasks.

Ghina in [14] proposed a model to predict the personalities of Twitter users with better performance than other prediction systems. It was done by an online poll utilizing the Big-five Inventory (BFI) survey, which was sent to 295 Twitter users and collected 511,617 tweets. The suggested model employed a Support Vector Machine (SVM) to test two alternative semantic approaches that combined SVM and BERT. The results reveal that combining these two approaches yields an accuracy score of 79.35 percent, with LIWC deployment improving the accuracy score by up to 80.07 percent.

Jianshan et al in [15] proposed a new way to formalize personality as a digital twin model by observing content postings and user preference behaviours. The deep neural multitask learning network model uses two forms of data representation to estimate a user's personality. Experiments suggest that integrating the two forms of data can increase personality prediction accuracy.

Junaid et al in [16] proposed a model to classify the input text into psychopathic and nonpsychopathic features. The majority of prior work on psychopath identification has been done in the realm of psychology, employing conventional methodologies like the SRP III method and small dataset sizes. As a result, they are motivated to create advanced computer models in the field of text analysis to detect psychopaths. This study investigates sophisticated deep learning approaches, such as attention-based BILSTMs, for expanding dataset size and effectively classifying input text into psychopath and nonpsychopath classes in order to discover psychopaths.

At the conclusion of this section, it must be affirmed that understanding personality is critical to target users with appropriate advertising, detect the dark side of the personality, and assist in the recruitment process, in which some companies require understanding personality prior to employment process acceptance in order to determine whether or not a candidate has aggressive behaviour. In this project, machine learning techniques will be utilized to create a model that predicts psychological personality scores based on user-generated content utilizing the Big-five model's multiple elements and metrics. The study's hurdles include gathering data and estimating a single user's precise score in each of the big-five factors.

IV. PROPOSED MODEL

Fig. 2 shows the suggested methodology. User-generated material (posts) is sent to the preparation phase, which includes data cleaning, data transformation, and text conversion to vectors so that the algorithms can develop a prediction model, manage missing values, and remove stop words. Model of this study has two main phases Machine learning phase and personality prediction phase. The dataset of users' posts contains two main characteristics (post and post category), as shown in table I. The features are then passed on to the machine learning methods phase, which uses 80% of the dataset for training and 20% for testing.

Then move to the personality prediction phase After building the classification model, which used to predict the category of the posts for each user. The result of this phase is a new csv file with post, predicted post category, and five new columns containing discrete values for each of the big-five traits for each post category; values from 0 to 5, which 0 means that a specific big-five trait does not meet that category and 5 means that a specific big-five trait strongly meets that category, a snap shot of the csv file is shown in Fig. 3.

Finally, the model can detect the user personality between the big-five by classifying his posts and then calculating the score as shown in Table II. Table II displays an example of personality prediction results for one user posts.
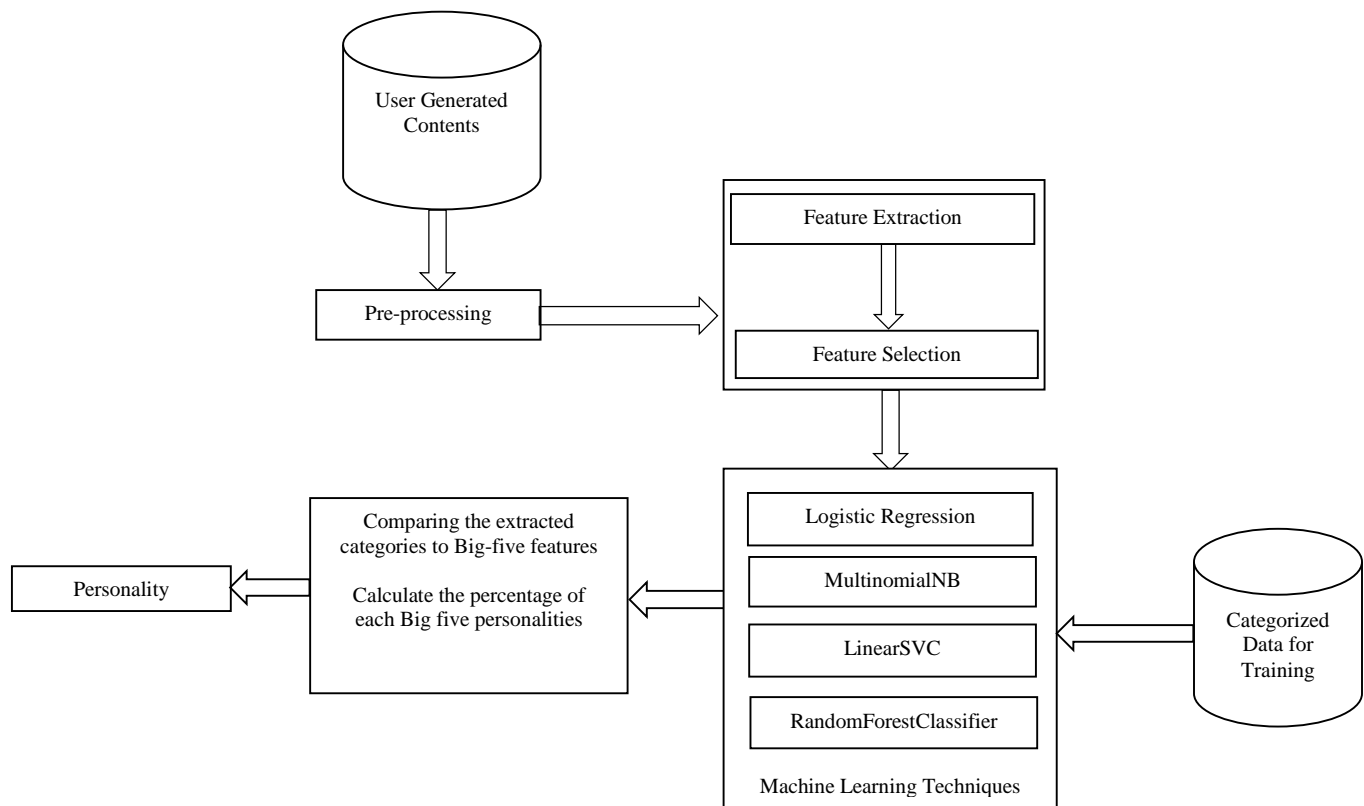
Fig. 2. The Proposed Model.

TABLE I. POST AND ITS CATEGORY

Post	Category
['Eman Tuhami: I swear if I said all the words of the world I cannot give you your place. During the four years college: with the feeling that I am your daughter happiness and happy day.\n Thousand million Congratulations my beloved makes us a good husband for you.']	['celebration']
['When you get paid for two months after suffering of waiting']	['Comic']

Post	Category	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
am the only one who was a h	['celebartion']	4	2	5	5	0
aid for two months after suff	['Comic']	2	1	5	3	0
['love and romantic']	['romantic']	5	3	4	3	0
I happy: O Lord: give you a rov	['celebartion']	4	2	5	5	0
u are good: Lord: and a thousa	['celebartion']	4	2	5	5	0
ives with each other\n Happy	['celebartion']	4	2	5	5	0
u are familiar with us you still	['celebartion']	4	2	5	5	0
ls\n I did not know what was	['happiness']	3	3	5	4	1

Fig. 3. CSV File with Big-Five Scores.

TABLE II. EXAMPLE OF PERSONALITY PREDICTION FOR ONE USER

Big-five Trait	Score
Openess to Experience	23.859386686611817 %
Conscientiousness	20.269259536275243 %
Extraversion	23.78459237097981 %
Agreeableness	23.036649214659686 %
Neuroticism	9.050112191473449 %

TABLE III.   MACHINE LEARNING ALGORITHMS ACCURACY

Model Name	Mean Accuracy
Linear SVC	89.37%
Logistic Regression	87.11%
Multinomial NB	78.81%
Random Forest Classifier	54.44 %

V.   EXPERIMENTAL RESULTS

Gathered dataset comes from the user's Facebook profile, post data, and medium post titles, and is based on a merged dataset. The final dataset comprises all of the Facebook postings as well as their categories. Dataset has 38 categories that assist in obtaining more precise findings. The total number of postings in the final dataset is 27,182. The categories are verified by the psychologist Dr. Noha Gamal.

Then the model applied the machine learning algorithms LR, MNB, LSVC, and RFC for predicting the post category for each post. The results shows that LSVC achieved the highest accuracy with 89.37% while the RFC got the lowest accuracy with 54.44%, the other results are shown in Table III, Fig. 4 and Fig. 5.

Fig. 4.   Comparison of Model Performance.

Fig. 5.   LSVC Confusion Matrix.

Our achieved accuracy proves the success of the proposed model by achieving more accuracy than previous work mentioned in the previous section.

The classification results are presented and explained in this section, and Table III shows a comparison of model performance. To forecast each user's personality traits, each user's postings are sent to the model as a distinct document, and the results identify the user's personality of the big-five.

VI. Conclusion and Future Work

Understanding personality qualities is vital for appropriately treating everyone based on his or her personality traits. The proposed model can be used to target users with appropriate advertising, detect the dark side of the personality, and aid in the recruitment process, in which some companies need to understand personality before accepting the employment process to know if this person has aggressive behaviors or not. The proposed model provesc the success of identifying user's personality according his actions and contents in social networks. The experimental results of this study represent an improvement because it can estimate the precise score of one user in each of the big-five factors. The proposed model, LinearSVC, has an accuracy of 89.37 percent. In future work, the proposed model is planned to use deep learning to improve accuracy.

References

[1] Cynthia CS Liem, Markus Langer, Andrew Demetriou, Annemarie MF Hiemstra, Achmadnoer Sukma Wicaksana, Marise Ph Born, and Cornelius J König, "Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. In Explainable and interpretable models in computer vision and machine learning", Springer, 2018, pages 197–253.

[2] Jang KL, Livesley WJ, and Vernon PA, "Heritability of the big five personality dimensions and their facets: a twin study", J Pers, 1996.

[3] Han, S., Huang, H., and Tang, Y., "Knowledge of words: an interpretable approach for personality recognition from social media. Knowledge-based systems", Elsevier, 2020.

[4] Schultz, D., and Schultz, S. E, "Psychology and work today: pearson new international edition coursesmart etextbook", Routledge, 2015.

[5] RA Power and M Pluess, "Heritability estimates of the Big Five personality traits based on common genetic variants", Translational Psychiatry, 2015.

[6] Ebru Gökaliler, Ozlem Alikilic, and Fırat Tufan, "Trust and Media: Reflection of the Big Five Factor Personality Traits on COVID-19 Pandemic Communication", Turkish Review of Communication Studies, 2022.

[7] Gozde Ikizer, Marta Kowal, Ilknur Dilekler Aldemir, Alma Jeftic, Aybegum Memisoglu-Sanli, Arooj Najmussaqib, David Lacko, Kristina Eichel, Fidan Turk, Stavroula Chrona, Oli Ahmed, Jesper Rasmussen, Raisa Kumaga, Muhammad Kamal Uddin, Vicenta Reynoso-Alcantara, Daniel Pankowski, and Tao Coll-Martín, "Big Five traits predict stress and loneliness during the COVID-19 pandemic: Evidence for the role of neuroticism", Elsevier,2022.

[8] Jose Ramon Saura, Domingo Ribeiro-Soriano, and Daniel Palacios-Marques, "From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets", Elsevier, 2021.

[9] Kamal El-Demerdash, Reda A. El-Khoribi, Mahmoud A. Ismail Shoman, and Sherif Abdou, "Deep learning based fusion strategies for personality prediction", Elsevier, 2022.

[10] Mr. R. Valanarasu, "Comparative Analysis for Personality Prediction by Digital Footprints in Social Media", Journal of Information Technology and Digital World, 2021, Pages: 77-91.

[11] Shakil Mahmud, Masud Rana, Fahim Rubaiyat Zahir, and Mohammad Rezwanul Huq, "Detection of Antisocial Personality Based on Social Media Data", Springer, 2021.

[12] Yang Li, Amirmohammad Kazameini, Yash Mehta, and Erik Cambria, "Multitask Learning for Emotion and Personality Detection", IEEE, 2021.

[13] Fatemeh Mohades Deilami, Hossein Sadr, and Mozhdeh Nazari, "Using Machine Learning-Based Models for Personality Recognition", arXiv, 2022.

[14] Ghina Dwi Salsabila and Erwin Budi Setiawan, "Semantic Approach for Big Five Personality Prediction on Twitter", Rumah Jurnal Elektronik Ikatan Ahli Informatika Indonesia, 2021.

[15] Jianshan Sun, Zhiqiang Tian, Yelin Fu, Jie Geng, and Chunli Liu, "Digital twins in human understanding: a deep learning-based method to recognize personality traits", International Journal of Computer Integrated Manufacturing, 2020.

[16] Junaid Asghar, Saima Akbar, Muhammad Zubair Asghar, Bashir Ahmad, Mabrook S. Al-Rakhami, and Abdu Gumaei, "Detection and Classification of Psychopathic Personality Trait from Social Media Text Using Deep Learning Model", Hindawi, 2021.