



Personality Prediction of Facebook Users Using Machine Learning

Razia Qamar*, Jasir Hassan †

Abstract: The usage of social networks is rising in the past few years. Most of the frequently used information is widely shared via social media networks including Facebook, Twitter, Instagram etc. Via these platforms, people can show their feelings, expressions, emotions, and opinions. These platforms provide meaningful content about users and their personality traits. Such information can be exploited to analyze the personality of the end-users. In this research, we explored a novel technique for personality modelling using Big 5 traits. We propose a comprehensive framework and machine learning pipeline that can be adopted for personality prediction. In order to solve the personality prediction problem, we proposed k-nearest neighbor (KNN) model along with several pre-processing steps.

Keywords: *Big Five traits, personality prediction, Facebook, my-personality, k-NN*

1. Introduction

Social media has become a vital part of everyday life during the previous two decades. It has been found that direct physical interaction among people is decreasing and social media interaction is taking dominance [2]. The reliance on social media platforms and keeping connected through them has grown rapidly, especially during the covid-19 epidemic [10]. Lockdown procedures are being used in a variety of fields, ranging from education to health care as well as entertainment. Facebook, Instagram, TikTok, Twitter, YouTube, WhatsApp and ResearchGate are all examples of social media platforms [9]. These networks daily generate enormous volume of data. Such large amounts of data have raised the interest of many experts, who are eager to anticipate people's behavior based on interaction data.

Predicting personality of an end-user based on social media data is a trending topic these days[1]. The topic has gained prominence specifically after the emergence of intelligent technologies. Some research studies have concentrated on predicting personality using different AI and ML algorithms based on social media data. A fundamental question is how do you define personality? A personality uniquely characterizes an individual [8]. [4] has defined personality as a combination of actions, thoughts, emotions that aims at describing

* Karachi Institute of Economics and Technology, Karachi

† Karachi Institute of Economics and Technology, Karachi

the various aspects of human behavior. According to [1], personality is a specific characteristic of a person or behavior that differentiates it from other users. Such researches that identify personality of a person based on its social media profile can become a prestigious scientific pursuit as they found a large number of useful applications. There have been conventional studies focused on identifying personality of an end-user that are based on human designed statistical features. The problem with these approaches are that they don't utilize the rich semantic information present in users' textual data [4].

Among the many studies in the literature, several studies found a significant relationship between personality and different Facebook activities. This includes comments on post, likes, hashtags, locations of post, mutual friends, events, and the posts. However, predicting personality of a person based on its social media interaction is quite challenging because of the diversity of online users [3]. This paper attempts to tackle this challenging problem by using machine learning. *Machine learning* is defined as the learning of patterns from the data that can be used to make future predictions. There are two major types of machine learning techniques: supervised learning and un-supervised learning. In *supervised learning*, the data along with its label is available and the machine learnings to find a mapping between data/features and its label [12]. In *un-supervised learning*, the label is not available. Generally, clustering and dimensionality reductions are performed using un-supervised learning [13].

Personality of a person has great impact on its life and achievements [1]. A personality not only defines person's actions and behaviors, but is also an indicator of person's happiness, preference, physical and mental health [4]. Companies that can utilize Facebook data to understand the personality of its customers may use this knowledge to develop their operations and reach their target market [1]. People with a high tendency to commit crimes may be readily anticipated using Facebook data, and people can learn about one other's personalities before entering into any connection with them [3]. The various fields that are specifically employing personality prediction are advertising, marketing, recruitment, and finances [9].

In this study we have employed the Big 5 personality model which is considered as the best model for describing the personality of person. We utilized publicly available Facebook's data to determine a person's personality based on the Big 5 personality model. [4], since an individual might have more than one personality feature. According to the Big 5 model, a person's personality varies in terms of five factors: openness, conscientiousness, extraversion, agreeableness, neuroticism. Big 5 is the most widely used model for personality prediction [8]. Each of these five personality qualities is associated with a classifier.

The objective of this research is to propose a novel k-nearest neighbor approach for predicting personality of end-user based on Big Five Personality model. Because of wide-scale usage of social media for interaction instead of physical means, it becomes quite difficult to identify the personality of an end-user [2].

It has been hypothesized that people spend a lot of time on social media, express their feelings, emotions, share thoughts, comments and likes the posts. These interactions can be used to analyze the personality of an end-user. However, it is quite challenging owing to diversity of users profile on social media. We proposed a comprehensive machine learning pipeline that can be used to identify the personality of an end-user. This study is directed towards analyzing the posts of one of the largest social media platforms i.e. Facebook. The user-base of Facebook is around 1.8 billion users with at least 800 million people spending at least 40 minutes daily on Facebook.

Rest of the sections of the paper has been organized as follows. We first discuss the relevant literature. This is followed by the proposed methodology. The results are presented afterwards. This is followed by conclusion and future work.

2. Related Work

There have been a lot of research work done on personality prediction. As we will discuss, personality prediction was done conventionally using questionnaires and interviews[8]. An overview of personality prediction using language features have been provided in [1]. [4] have classified the literature on personality prediction as those based on machine learning and deep learning techniques.

According to various studies, machine learning has been found to provide very good results for personality prediction [8]. [4] discussed various approaches to machine learning based personality prediction such as those based on support vector machines, Gaussian approaches. Then, they have also discussed approaches based on long short term memory (LSTM), gated recurrent unit (GRU), and convolutional neural network (CNN). LSTM solves the problem of RNN by maintaining long-term dependencies. It has been used in NLP for a long time. CNN comprises layers such as convolution, maxpooling. It has been used actually for image processing. But some researchers have used CNN for NLP tasks as well. In the paper, the authors, then proposed a CNN based approach for personality prediction that also used AdaBoost for optimize learning rate while training.

The authors in [2] have used Big 5 Personality traits and various deep learning techniques for personality prediction. They achieved an accuracy of around 74.17%. As far as conventional approaches are concerned, personality assessments were performed using interview and self-report inventories [1]. They were performed by domain experts and quite time consuming. A multi-objective model based on domain knowledge to predict personality has been proposed in [8].

Personality prediction found its application not only in psychology but in computer science as discussed in introduction. However, owing to the diversity of the online user, it is quite challenging to predict personality of an end-user based on its social media interaction. In this

direction, [3] proposed a novel approach i.e. a personality prediction framework. It comprises outlier elimination, training data set selection, and personality prediction. [19] argued that social media is a source for personality prediction.

Besides Big Five, other models that have been used for personality prediction are MBTI [9] with some studies comparing both of them. In [9], a dataset labelled with MBTI personalities were used and several machine learning algorithms were compared such as Naïve Bayes, Logistic Regression and artificial neural networks. They also tested several clustering oriented solutions and achieved an accuracy of 90%.

Several machine learning techniques have been implemented in [10] where Word2Vec model have been used for text representation and synthetic minority oversampling techniques have been used to address data imbalance. All of the above-mentioned studies used social media in English to predict personality traits such as based on the Big Five Personality types.

A recent study was undertaken to develop a personality prediction system in Bahasa Malaysia using Twitter, based on the Big-Five personality models. In [5], relation between psychological stress and personality has been discussed.

Deep learning has been used in various studies on personality prediction as discussed in previous paragraph. Classification method Big Five Social Media Personality Models Facebook has been the dominant approach [1][5]. In this direction, combining deep learning and Big Five Personality trait model was done in [5] where neural network has been used. The most important aspect of this research is the use of novel features such as likes, events, groups, tags, updates, network size, relationship status, gender and age.

In another direction, several swarm intelligence optimization techniques such as Grey Wolf is used for hyper parameter optimization with various machine learning models for personality prediction [11].

As far as personality prediction is concerned, there are various personality prediction models available in literature [2]. This includes Big Five Personality, MBTI (MyersBriggs Type Indicator) or DISC (Dominance Influence Steadiness Conscientiousness) [10]. Deep learning has been used for personality prediction in [9].

A systematic literature review on personality prediction has been done in [15]. A review of data mining based approaches for personality prediction has been performed in [29]. A transformer based model XLNet has been used for personality prediction in [16].

Personality prediction based on human-machine dialogue has been performed in [17]. A meta-analysis of five-factor model of personality and social media has been performed in [18]. [20] has proposed a humanoid system for personality assessment based on Big-5 model.

Chatbot personality using conversation cues in social media messages have been performed in [21]. Mental health prediction based on social media data has been performed in [23].

An ensemble of machine learning algorithms have been proposed in [27] for personality prediction. A stacked ensemble of machine learning classifiers are used based on semantic and lexical features. In addition, several deep learning algorithms are also used. A multimodal approach to personality prediction that combines visual and audio features have been proposed in [30]. ChaLearn-16 dataset has been used for evaluation.

Recently, large language model(LLM) has gained a lot of popularity. LLM based approaches for personality prediction has been used in [28]. In a number of studies, personality prediction for animals have been performed in [24, 25, 26]. In another direction, consumer behavior was analyzed using a mathematical model based on social media data. Table 1 summarizes the literature.

Table 1: Summary of selected literature on personality prediction

Reference	Contribution
Joshanloo [5]	The relation between stress and personality trait is discussed
Jayaraman et al. [16]	A transformer-based model has been used for personality prediction
Guo et al. [17]	Personality prediction for human-machine dialogue has been performed
Jaffar et al. [20]	Humanoid for personality prediction
Heppner [21]	Analysis on chatbot data has been performed
Amirhosseini et al. [24]	Studies on animal personality prediction
Guerrero [27]	An ensemble of machine learning classifiers has been used for personality prediction
Hu et al. [28]	LLM-based approach for personality prediction

Based on the studies, it can be concluded that there has been a lot of interest in personality prediction. However, very few approaches have tried to develop a comprehensive pipeline for personality prediction. The next section presents a pipeline for personality prediction.

3. Proposed Approach

We now present our proposed approach based on Big 5 personality. The first step is the data acquisition which is done via scrapping/ downloading/ sourcing the dataset from Kaggle. In the next step, data normalization is performed. Then, k-fold cross validation is performed. We

then split the dataset into training and testing. Finally, k-nearest neighbor algorithm is used for classification of personality traits.

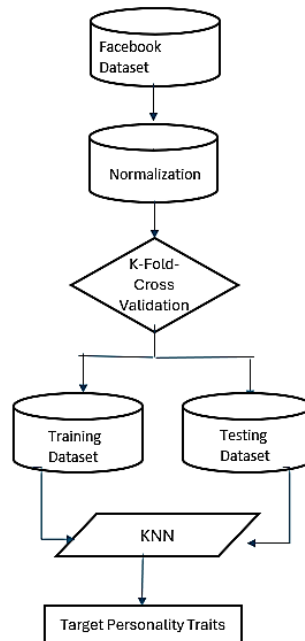


Figure 1: Flowchart of proposed methodology

3.1. The Processing pipeline

The processing pipeline outlines various steps such as data acquisition, preprocessing, normalization etc. Following are the major steps performed as shown in Figure 1:

- a. The first step is the data acquisition which is performed via downloading the data from Kaggle
- b. The preprocessing step removes noise from the data such as URLs, symbols, name etc. These URLs / symbols don't add any value to the text while used for analyzing the end-users personality
- c. Then, case normalization is done. All the text is converted to lower case. This reduces the vocabulary size while analyzing the text.
- d. Stop words are removed. Stop words are commonly used words such as the, a, an, there etc. These words are common across various text and doesn't add any

discriminative capability to the texts.

- e. Stemming/ lemmatization is performed to identify the root words. This helps in reducing the size of the vocabulary.
- f. Feature selection is performed afterwards.
- g. Classification of the personality is performed



Figure 2: Processing pipeline for the proposed approach

3.2. Personality Measurement Models

Personality can be described by various theoretical perspectives and various levels of abstractions. An example of such an abstraction, is the personality trait, the notable one is the Big 5 Personality Trait / dimension [6]. According to this study, there are five major attributes that describe personality of a person known as “Big 5”. They are well documented for a personality and has been used by researchers in literature. These attributes are openness, conscientiousness, extroversion, agreeableness, and neuroticism as shown in Table 2. According to the literature, Big Five Factor is used to explain personality of a person. This model was proposed in 1990 and is still in use.

Table 2. Big 5 Personality Traits [6]

Openness	Conscientious	Extroversion	Agreeableness	Neuroticism
Imaginative, Wide interest, Curious, Intelligent, Artistic, Unconventional	Organized, Disciplined, Planner, Goal oriented, not impulsive	Energetic, Forceful, Adventurous, Enthusiastic	Sympathetic, Straight forward, Compliance, Generous	Anxious, Tense, Worried, irritable, impulsive, shy

Besides Big 5, MBTI is another popular personality measurement model [10]. It describes a person's preference along following dimensions: Introvert/Extrovert, Intuition/Sensing, Thinking/Feeling, Judgment/ Perception. These attributes can be consolidated to create a spectrum of 16 different personality types. MBTI has been used in career development, counselling and relationship improvement. Another popular model is DISC [10].

However, according to researchers, Big 5 is more comprehensive in assessing an individual's personality as compared to MBTI and DISC [10].

3.3. Dataset

The dataset has been sourced from Kaggle named as “myPersonality”. In this dataset, there are a total of 250 records of social media users and almost 10,000 statuses with user id shown in Table 3 [7].

Table 3: Snapshot of the dataset

	#AUTHID	STATUS	sEXT	sNEU	sAGR	sCON	sOPN	cEXT	cNEU	cAGR	...
0	b7b7764cfa1c523e4e93ab2a79a946c4	likes the sound of thunder.	2.65	3.00	3.15	3.25	4.40	n	y	n	...
1	b7b7764cfa1c523e4e93ab2a79a946c4	is so sleepy it's not even funny that's she ca...	2.65	3.00	3.15	3.25	4.40	n	y	n	...
2	b7b7764cfa1c523e4e93ab2a79a946c4	is sore and wants the knot of muscles at the b...	2.65	3.00	3.15	3.25	4.40	n	y	n	...
3	b7b7764cfa1c523e4e93ab2a79a946c4	likes how the day sounds in this new era...	2.65	3.00	3.15	3.25	4.40	n	y	n	...

3.4. Feature selection

The next step is the selection of the features. We considered two columns from above dataset: one is *STATUS* and the second column is *LABEL*. We used Big 5 personality traits and calculated the Big 5 personality distribution manually. Table 4 represents the distribution [5]. Each of these columns represents the big 5 personality traits.

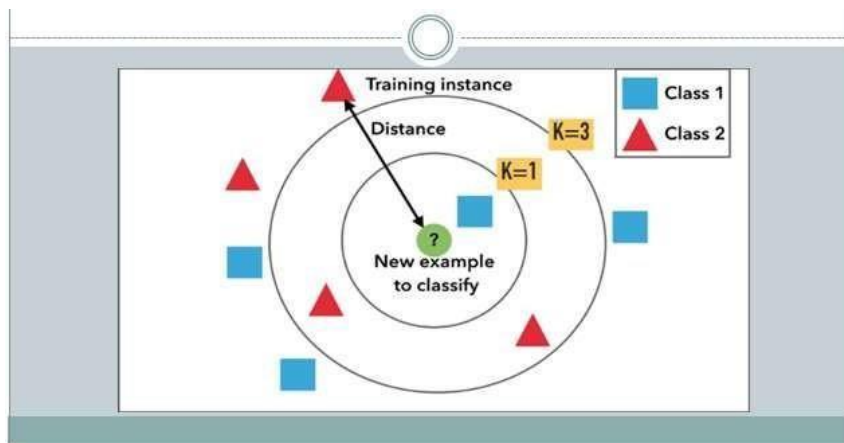
Table 4. Big 5 Personality Distribution.

Value	OPE	NEU	AGR	CON	EXT
Yes	7370	3717	5268	4556	4210
No	2547	6200	4649	5361	5707

3.5. Model classification

As discussed, machine learning and deep learning have been employed in several studies for the classification of the data. Amongst the traditional machine learning algorithms are Naive Bayes, decision trees, neural network, support vector machine, Logistic Regression, gradient boosting. For this study, the paper utilized k-nearest neighbor. For validation of the model, we employed 0-fold cross validation. In other words, data is divided into 10-folds. One is used for validation and rest for validation in each pass [2].

We used k-NN (k-Nearest Neighbour) for personality prediction using user's status from myPersonality dataset. The Python library Keras/ tensorflow is used. Various pre-processing steps such as cleaning/ vectorization of the text sample into a 2D integer tensor is performed before classification. The dataset is distributed in two parts: one for training and second for testing with the distribution of 66% and 33%. The 66% of the data is used for training and 33% for testing. We made one column for label named *OUTPUT* that will show the traits in which personality belong. Figure 3 shows the k-nearest neighbor algorithm that we have used.

**Figure 3: k-nearest neighbor algorithm**

3.6. Mathematical Formalism

Personality prediction can be considered a machine learning problem. The goal is to determine the personality y based on a set of features $\mathbf{x}$. We can express this as:

$$y = f(\mathbf{x}; \theta) + \epsilon$$

Here, $\mathbf{x}$ is the input feature vector. Y is the personality trait (can be a vector). f is the predictive function parameterized by θ . ϵ is the error term for noise. For personality prediction using social media content, the features can be lexical features such as word frequency, TF-IDF; sentiment based features such as polarity, motion score; linguistic style feature such as use of first/second or third person. In addition the use of hashtags and emojis also have significance.

k-nearest neighbor is primarily a non-parametric method. It can be used for both classification and regression. Suppose, we have a set of data points with feature $\mathbf{x} \in \mathbb{R}^d$. y_i is class label. Then compute distance between $\mathbf{x}_q$ and all data points $\mathbf{x}_i \in D$.

$$d(\mathbf{x}_q, \mathbf{x}_i) = |\mathbf{x}_q - \mathbf{x}_i|^2$$

Then, assign label categories based on majority voting among k-neighbors as follows:

$$y_q = \arg \max_c \sum_{i \in N} I(y_i = c)$$

3.7. Train / test split

We split data into two parts training and testing training data contain 66%. We use 33% data for testing. The model performs quite well with the test data and shows good generalizability.

```
[ 'likes the sound of thunder.'
'is so sleepy it's not even funny that's she can't get to sleep.'
'is sore and wants the knot of muscles at the base of her neck to stop hurting. On the other hand, YAY I'M IN ILLINOIS! <3'
'likes how the day sounds in this new song.' 'is home. <3'
'www.thejokerblogs.com'
'saw a nun zombie, and liked it. Also, *PROPNAM* + Tentacle!Man + Psychic Powers = GREAT Party.'
'is in Kentucky. 421 miles into her 1100 mile journey home.'
'was about to finish a digital painting before her tablet went haywire. Is now contemplating the many ways she wishes to exact her revenge on faulty technology.'
'is celebrating her new haircut by listening to swinger music and generally looking like a doofus.'
'has a crush on the Green Lantern.' 'has magic on the brain.'
'saw Transformers, Up, and Year One this week. Good movie overload. :D'
'who wants to meet up on schedule pick-up day at Oviedo?'
'desires the thrill of inspiration. Also, money.'
'is going to bed at 9:30! Yeah!'
'is reading, admiring her permit, and occasionally glancing at her ner McDonald's uniform.'
'thinks intangibility should be an option in reality settings.'
'is tired. *PROPNAM*, let me go to sleep pløx.'
'is discovering the many flavors of insomnia.'
```

a. Training data

```

["It's not premarital sex if no-one will marry you..."
'Mmm... Damn new fucking children only watching blonde people fuck.'
'Fuck.... Laptop sleeps retardedly slop whjen i'm wasted.'"
'Leprechaun Pikachu!!' 'OMFG *PROPNANE* is in Elf...'
'Sand is overrated. It's just tiny, little rocks...."
'Whooop! Runshaw's Off. To the Sledging..."
'Those who died are justified, for wearing the badge, they're the chosen whites. You justify those that died by wearing the badge, they're the chosen whites"
'Way runshaw out tomorrow as well.'
'Fuck you Runshaw, I want to sleep tomorrow.'
'and *PROPNANE*. Leaders of an Awesome (yet possibly painful) New World Order.'
'"I\'m Covered in Bees!"'
'I want to reconcile the violence in your heart. I want to recognize your beauty is not just a mask. I want to exorcise the demons from your past. I want to satisfy the undisclosed desires in your heart. ?'
'I was right. Glee has ruined Don't Stop Believin'."
'I dream of a better tomorrow, where chickens can cross the road and not be questioned about their motives.'
'What? \x85 The Carthaginians are attacking? God, I knew they'd do that. What? \x85 They are attacking over the Alps? Damn, I knew they'd do that. What? \x85 They're coming on elephants? \x85 Where'd they get the elephants? There aren't any elephant s in Europe."

```

b. Testing data

Figure 4: Training/testing data split

4. Results

We now present the results. Figure 5 shows the results. The results show which status belong to which type of personality such as openness, conscientious, extroversion, agreeableness, and neuroticism. Some status belong to two or more types. Table 4 also shows the results.

Table 4: Results of proposed approach

	Accuracy (%)	Precision	Recall	F1
Logistic Regression	83	82	81	82
Decision Trees	84	83	80	84
SVM	85	78	80	85
Proposed approach	87	86	85	86.5

Figure 5 shows the results of the implementation. We have compared the results with Logistic regression, decision trees, SVM and proposed approach. It can be seen that logistic regression provides 83% accuracy, while decision trees and SVM provides accuracy of 84 and 85% accuracy. However, proposed approach outperforms all of them with an accuracy of 87%. Figure 6 provides the detailed comparison of results.

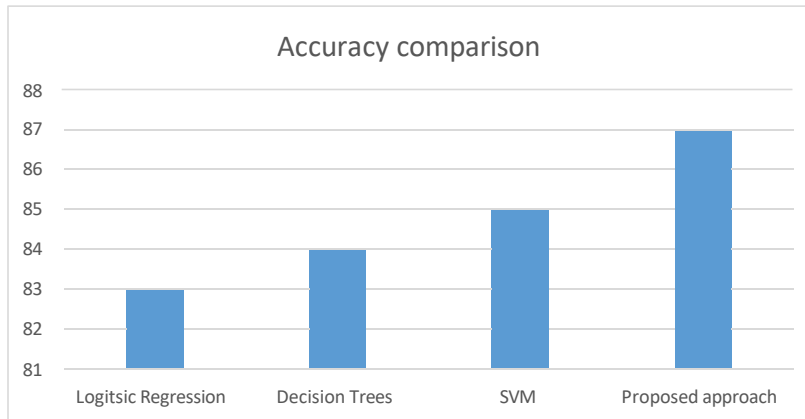


Figure 5: Accuracy comparison of proposed approaches with SOTA

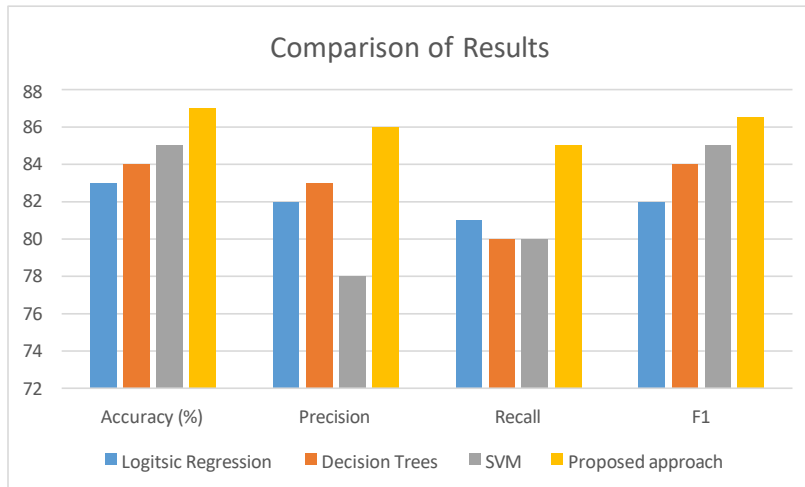


Figure 6: Comparison of Results for proposed approach and other approaches

5. Conclusion

Personality prediction has been a topic of intense research. It predicts a user's personality traits from the information available on social media platforms. For instance, [1] has identified that personality prediction can have benefit in social network analysis, product recommendations, deception detection, authorship attribution, sentiment and opinion mining. In this study, we have used k-nearest neighbor algorithm for personality prediction. It is also found during the research that using machine learning and deep learning algorithms can outperform the results of previous studies that are using classical algorithms.

Therefore, in future, we plan to augment the dataset with more information. We also plan to use Fuzzy Logic and Genetic Algorithm and other architectures, and other processes to improve this prediction system.

This study explored the performance of kNN for personality prediction. It employed the big five personality based on the publicly available Facebook data. This study proposes a classification model for Big 5 using kNN. Future work can be done on improving the accuracy of the model via acquiring more data and performing k-fold cross validation of the data available on other social media platforms besides Facebook.

6. Future Work and Implications

The proposed work can be extended in various directions. In the future work, we will prove that with Fuzzy Logic and choosing machine-learning algorithm for better accuracy and transfer the proposed framework from the same domain dataset to the different domain dataset.

In addition, will analyze with more data and extensive k-fold cross validation. Besides, exploration of personality prediction beyond Facebook such as Twitter, LinkedIn can be performed and is left as a future work. For instance, personality prediction on WhatsApp user has been performed using deep learning in [22].

The problem with machine learning based approaches is that it is very difficult to extract the right kind of features required for prediction[4], hence, future work can also be done on using deep learning based approaches for personality predictions.

The implications of this study are very broad. It can be used for business and marketing such as for personalized advertising, customer service optimization. It can be used for human resource and recruitment, mental health and well-being. There are various social and relationship applications. However, there are several limitations such as privacy and data misuse, bias and discrimination, psychological manipulation.

References

- [1] Aung, Z.M.M. and Myint, P.H., 2019, July. Personality prediction based on content of Facebook users: A literature review. In *2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)* (pp. 34-38). IEEE.
- [2] Tandra, T., Suhartono, D., Wongso, R. and Prasetyo, Y.L., 2017. Personality prediction system from facebook users. *Procedia computer science*, 116, pp.604-611.

- [3] Zhong, X.F., Guo, S.Z., Gao, L., Shan, H. and Xue, D., 2018, February. A general personality prediction framework based on facebook profiles. In *Proceedings of the 2018 10th International Conference on Machine Learning and Computing* (pp. 269-275).
- [4] Deilami, F.M., Sadr, H. and Nazari, M., 2022. Using machine learning based models for personality recognition. *arXiv preprint arXiv:2201.06248*.
- [5] Joshanloo, M., 2024. Longitudinal relationship between psychological distress and personality traits. *Stress and Health*, 40(2), p.e3305.
- [6] John, O.P., 1999. The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research/Guilford*.
- [7] <https://www.kaggle.com/code/testaccount93/mypersonality/data>, last accessed on December, 2024
- [8] Han, N., Li, S., Huang, F., Wen, Y., Su, Y., Li, L., Liu, X. and Zhu, T., 2023. How social media expression can reveal personality. *Frontiers in Psychiatry*, 14, p.1052844.
- [9] Abdurahman, S., Vu, H., Zou, W., Ungar, L. and Bhatia, S., 2024. A deep learning approach to personality assessment: Generalizing across items and expanding the reach of survey-based research. *Journal of Personality and Social Psychology*, 126(2), p.312.
- [10] Ryan, G., Katarina, P. and Suhartono, D., 2023. Mbt personality prediction using machine learning and smote for balancing data based on statement sentences. *Information*, 14(4), p.217.
- [11] Talaat, F.M., El-Gendy, E.M., Saafan, M.M. and Gamel, S.A., 2023. Utilizing social media and machine learning for personality and emotion recognition using PERS. *Neural Computing and Applications*, 35(33), pp.23927-23941.
- [12] Rajab, A., Farman, H., Islam, N., Syed, D., Elmagzoub, M.A., Shaikh, A., Akram, M. and Alrizq, M., 2023. Flood Forecasting by Using Machine Learning: A Study Leveraging Historic Climatic Records of Bangladesh. *Water*, 15(22), p.3970.
- [13] Muhammad, G., Saeed, U., Islam, N., Kumar, K., Hussain, F., Khurro, M.A., Shaikh, A.A. and Ali, I., 2022. Gvdeepnet: Unsupervised deep learning techniques for effective genetic variant classification. *Pakistan Journal of Engineering and Technology*, 5(1), pp.16- 22.
- [14] Islam, N., Mehboob, M. and Javed, R., 2023. Analytics of User Behaviors on Twitter Using Machine Learning. In *Handbook of Research on Artificial Intelligence Applications in Literary Works and Social Media* (pp. 302-337). IGI Global.

- [15] Singh, S. and Singh, W., 2024. AI-based personality prediction for human well-being from text data: a systematic review. *Multimedia Tools and Applications*, 83(15), pp.46325-46368.
- [16] Jayaraman, A.K., Ananthakrishnan, G., Trueman, T.E. and Cambria, E., 2024. Text-based personality prediction using XLNet. In *Advances in Computers* (Vol. 132, pp. 49-65). Elsevier.
- [17] Guo, A., Hirai, R., Ohashi, A., Chiba, Y., Tsunomori, Y. and Higashinaka, R., 2024. Personality prediction from task-oriented and open-domain human-machine dialogues. *Scientific Reports*, 14(1), p.3868.
- [18] Meynadier, J., Malouff, J.M., Schutte, N.S. and Loi, N.M., 2024. Meta-analysis of associations between five-factor personality traits and problematic social media use. *Current Psychology*, 43(27), pp.23016-23035.
- [19] Dandannavar, P.S., Mangalwede, S.R. and Kulkarni, P.M., 2018, December. Social media text-a source for personality prediction. In *2018 international conference on computational techniques, electronics and mechanical systems (CTEMS)* (pp. 62-65). IEEE.
- [20] Jaffar, A., Ali, S., Iqbal, K.F., Ayaz, Y., Ansari, A.R., Fayyaz, M.A. and Nawaz, R., 2024. A comprehensive multimodal humanoid system for personality assessment based on the big five model. *IEEE Access*.
- [21] Heppner, H., Schiffhauer, B. and Seelmeyer, U., 2024. Conveying chatbot personality through conversational cues in social media messages. *Computers in Human Behavior: Artificial Humans*, 2(1), p.100044.
- [22] Saeidi, S., 2024. Identifying personality traits of WhatsApp users based on frequently used emojis using deep learning. *Multimedia Tools and Applications*, 83(5), pp.13873-13886.
- [23] Fudholi, D.H., 2024. Mental health prediction model on social media data using CNN-BILSTM. *Kinetik: Game technology, information system, computer network, computing, electronics, and control*, pp.29-44.
- [24] Amirhosseini, M.H., Yadav, V., Serpell, J.A., Pettigrew, P. and Kain, P., 2024. An artificial intelligence approach to predicting personality types in dogs. *Scientific Reports*, 14(1), p.2404.
- [25] Liang, Q., Liu, D., Zhang, D., Wang, X., Zhu, B. and Wang, F., 2024. Machine learning-based aggressiveness assessment model construction for crabs: A case study of swimming crab

Portunus trituberculatus. *Aquaculture*, 593, p.741304.

[26] Kakani, V., Kesani, B., Thotakura, N., Bodapati, J.D. and Yenduri, L.K., 2024, April. Decoding Animal Emotions: Predicting Reactions with Deep Learning for Enhanced Understanding. In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)* (pp. 1-6). IEEE.

[27] Serrano-Guerrero, J., Alshouha, B., Bani-Doumi, M., Chiclana, F., Romero, F.P. and Olivas, J.A., 2024. Combining machine learning algorithms for personality trait prediction. *Egyptian Informatics Journal*, 25, p.100439.

[28] Hu, L., He, H., Wang, D., Zhao, Z., Shao, Y. and Nie, L., 2024, March. Llm vs small model? large language model based text augmentation enhanced personality detection model. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 16, pp. 18234-18242).

[29] Gardade, M.R. and Khambre, P., 2021. A REVIEW ON PERSONALITY PREDICTION VIA SOCIAL MEDIA USING DATA MINING. *INTERNATIONAL JOURNAL*, 6(1).

[30] Suman, C., Saha, S., Gupta, A., Pandey, S.K. and Bhattacharyya, P., 2022. A multi-modal personality prediction system. *Knowledge-Based Systems*, 236, p.107715.

[31] Chaudhary, K., Alam, M., Al-Rakhami, M.S. and Gumaei, A., 2021. Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics. *Journal of Big data*, 8(1), p.73.