

User Personality Prediction on social media : An Overview

Poonam L Patil¹, Dr. Satish R. Jadhao²

¹Dept of computer engineering Gokhale Education Society's R. H. Sapat College of Engineering Management Studies and Research, Nashik-5, Maharashtra, India.

²Dept of computer engineering, Assistant Professor, Gokhale Education Society's R. H. Sapat College of Engineering Management Studies and Research, Nashik-5, Maharashtra, India.

Abstract - User Personality is a combination of Some attributes like behavioral, temperamental, emotional, and mental that characterizes a unique individual. Recently, predicting personality from the social media to be lean research area in computational linguistic. To establish the personalities of people has always been of extreme significance to the researcher's attention. With increasing development of social networks, there are many approaches to have been developed to find user personality on social media depending on social activities. With the development of Internet usage, people can communicate their opinion, activities, feeling and thoughts through social media. Posts, comments and status updates made by users of social media can reveal. personal information. The main objective of this is to review the work carried out for personality prediction on social media using the big five model. This paper may develop a support material for those who want to know about the personality prediction form the social media update content. We concern that our present, specifically in epitomizing the earlier discovery and in analyzing the directions for future research in this area.

Key Words: :Social media, Big five model, personality prediction.

1.INTRODUCTION

Now a Days, Social Media such as Facebook, Twitter become most popular for internet users. The number of the users increases day by day on the social media. Recently, social networking sites become an outstanding podium for web surfers to post the opinion, feelings, comment and updates [1,2,3]. The Fig 1. shows that the number of monthly social media users in India from Jan 2020- June 2020. The activity of user on these social media is important to find and understand the online behaviors and personality of that user. Personality embraces mood, attitudes and opinion also, the components, incorporating the applicable criteria that follow, interaction with other people. It includes the behavioral characteristics that distinguish one person with another. The personality of person is corresponding to the status profiles on social media. Although psychology experts can observe personality traits from social medial, it is very costly and required more

time. The classical approach of personality prediction through personality assessment by using survey or questionnaires are costly and required more time [4]. To solve this issue, many computational linguistic techniques have been used to describe individual's personality from social media data.

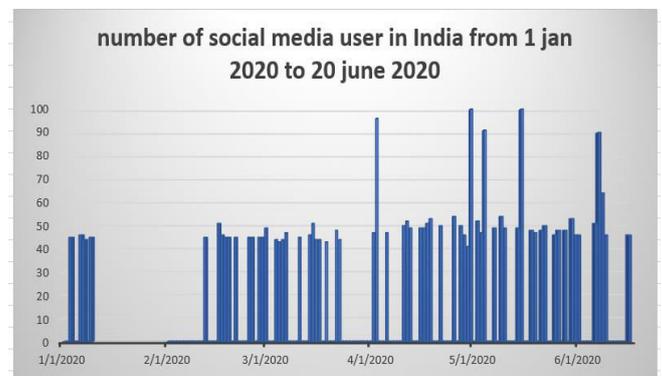


Fig 1 number of social media users in India from 1 January 2020 to 20 June 2020 (using google trends)

A personality prediction model based on texts extracted from social media that can be useful in several areas, including marketing intelligence and social psychology, due to the high volume of information generated and the exposure level. The recognition of personality traits helps to find out the mutual conduct and may provide a subjective view to text mining in social media, such as: sentiment analysis, text clustering, and recommendation systems. There have been several approaches to automatically predict users' personality based on different kind of datasets such as essays, video and social media post, and social media behavior [5].

This Paper scrutinizes the different models used for predicting the personality of user on various social media depends on their language used on social media status. The user action on social networks keeps a great platform for researchers to study and understand their online behaviors, preferences and personalities. Different personalities are related to the evolution of different social relationships and cooperation behaviors on status profiles. The main objective of this paper is to give a brief of the development of personality prediction from text on social media. This paper also aims to contribute an describe the issues that are realized in personality prediction.

This document shows the suggested format and appearance of a manuscript prepared for SPIE journals. Accepted papers will be professionally typeset. This template is intended to be a tool to improve manuscript clarity for the reviewers. The final layout of the typeset paper will not match this template layout.

1.1 Big Five Model

Big five model is the most used recently to measure of personality structure. The big five personality model identifies five types of personalities and every individual fall into at least one of these types. This model is also called OCEAN model. the models describe the five domain of personality Openness, Conscientiousness, extroversion, Agreeableness, and Neuroticism [6,7,8]. This is suitable model among the language and computer science researchers, and it has been used as a structure for both personality traits recognition and simulations. The Table 1 shows the overview of Big-Five personality traits and their characteristics.

Personality Traits	Some characteristics
Openness(O)	Curious, open to new ideas, creative, intellectual, original, Artistic, curious, imaginative, curious, intelligent, and imaginative
Conscientiousness(C)	Efficient, organized, responsible, organized, and persevering. Conscientious individuals are extremely reliable and tend to be high achievers, hard workers, and planners.
Extraversion(E)	Outgoing, talkative, Express positive emotion, satisfied, Friendly, Energetic, active, assertive, outgoing, amicable, assertive, Friendly and energetic, extroverts draw inspiration from social situations.
Agreeableness(A)	Affable, tolerant, sensitive, trusting, kind, and warm truthful, helpful, nurturing, optimistic, concerned, trusting of others, cooperative, compassionate, nurturing.
Neuroticism(N)	Anxious, Irritable, temperamental, moody, angry, upset tense, self-pitying, insecure, sensitive. Neurotics are experiencing negative emotions.

Table-1: Overview of Big Five Model

2. PERSONALITY PREDICTION MODELS

Personality prediction is finding the personality traits based on Big Five model. Recently the Big five model mostly used by researches to predict the professional behaviors and the

relationship on social media. Researches pointed a relationship between personality and the use of social media, like Facebook and Twitter. personality recognition in terms of analyzing user activity within Facebook or any other social media, we collected information about the personality traits of users and their profiles on social media [9]. The different machine learning algorithms are available to predict the personality of user. Table 3 shows the overview of the personality prediction model used.

A. Naïve Bayesian Classifier

Z Wei et al. [10] Predicted the Personality using web searching result. They use two step feature selection strategy. The Proposed model uses Naïve Bayesian classifier for prediction. The limitation or future scope of this system is to improve the performance of the system. To design semi supervised MLNP classifier to design the algorithm.

N.M.A Lestari et al. [11] For the Training data the Naïve Bayes method is best method in machine learning. In which Naïve Bayes method is used to determine the user’s personality from the written text. The Proposed model uses the preprocessing, which is done by case folding, Tokenizing Filtering, Stemming. After the preprocessing they perform the classification with Naïve Bayes. The Naïve Bayes method consist of two phases, learning phase and classification phase. user write its self-descriptive text that is helping to predict the user personality then match them to find the partner on online dating sites.

Ana C.E.S Lim et al. [12] have proposed personality traits prediction in text groups and extended the problem of personality prediction into a multi label classification problem. This is a now a advance as an individual may possess more than one personality traits. The authors used the Naïve Bayes Algorithm to analyze tweets and named their model as Bayesian Personality Predictor. Their approach was divided into three steps namely, preprocessing, transformation and classification, where in preprocessing certain attributes were extracted from the tweets and then in the second phase multi label sets were mapped into five single label training sets. Finally, in the third phase semi supervised classification takes place with the help of these training sets and meta-attributes as input to the classifier. The algorithm was evaluated with k-fold cross validation and metrics Accuracy, Recall and Precision. Brazilian TV shows were used as a benchmarking for personality analysis tool.

B. KNN

G Farnadi, S Zoghbi et al. [13] they extracted linguistic features, social network features, time-related and others feature from 255 Facebook users, and adopted KNN, Naïve Bayes and SVM to predict personality. KNN is received highest precision accuracy (0.54%) based on all of these features while it is reached to the highest precision (0.71%) by using social network features.

Arifin, K . E. Purnaha [14] have proposed personality to classify text in Indonesian language into emotion expression classes. The text weighing was done by using the TF-IDF

method (Term frequency and Inverse Document Frequency). In the classification process, KNN was used this method could classify the emotion expression of Indonesian language. By this method the Indonesian text classification in basic emotion classes. In Which they predict the Basic features related to emotions, such as sadness, surprise, Fear, Joy, anger. The KNN shows the best accuracy.

Ryan T, Xenos S [15], Method Evaluate the communication between users from a different perspective, they try to inspect into who uses Facebook and the relationship between the Big Five personality traits, presence of shyness, selfishness, aloneness, and Facebook usage. The results show that the Facebook users are more extraverted, But the non-permanent users and feeling socially alone than non-users. Also, the demand of Facebook movement with distinct structures are observed to vary that are results of certain characteristics, such as egoism, feeling alone and modesty. The other research study has used the data based on personality to examine the relationship between various types of Twitter users and personality, including popular users and famous ones [16]. This study has possessed just 335 users that shows the Twitter accounts in their Facebook profiles.

C. Support Vector Machine

Iacobelli, F et al. [17] developed a supervised personality detection model for Modern Greek with linguistic features and psychological features. They use the dataset corpus of blogs to compare classification feature selection also they show the results to find the information related to the personality of user. The various feature construction setting, such as stemmed bigrams, no exclusion of stop word the Boolean. The SVM classifier used for building the machine learning model, they demonstrated that personality and language can be successfully ported from English to other languages. The result shows that the SVM gives best accuracy range from 84.36 % to 70.51%.

Soujanya Poria et al. [18], authors proposed a new approach for personality detection which is based on incorporating the sentiment, affective and common-sense knowledge from the text using resources. In their approach, they combined common sense knowledge-based features with phyco-linguistic features and frequency-based features and later the features were employed in supervised classifiers. Further, they developed five support vector machine models for five personality traits. They designed five Social Media Optimization (SMO) based supervised classifiers for five personality traits. Their experimental results show that the use of common-sense knowledge with affective and sentiment information enhances the accuracy of the existing frameworks which use only psycho-linguistic features and frequency-based analysis at lexical level.

Y. Liu et.al built [19] they evaluated the Support Vector Regression (SVR) and Latent Dirichlet Allocation (LDA) models to recognize Big Five personality traits using

MyPersonality Facebook user's dataset. Their results proved that LDA models achieve better performance than the SVR model. Furthermore, LDA increases computational efficiency up to 64%.

K. Peng, L. Liou, C. Chang and D. Lee [20] They recognize the personality based on historical action logs on social media. They use the Chinese text for predicting the personality. They collected the dataset with post and personality score of 222 Facebook users. They use the Jieba segmentation tool for data preprocessing and support vector machine (SVM) algorithm for the personality classification. The result shows that the precision and recall can be improved with the help of text segmentation.

T. Tandera, D. Suhartono et al. [22] They uses the deep learning approach for personality prediction. This implementation method gives the analysis of various machine learning approaches for personality prediction they use the mypersonality facebook user dataset for comparative study. The deep learning approach gives the average accuracy of 74.17%.

D. boosting

Alireza Souri et al [21] their study on recognize the personality of the user activity on Facebook social media. They collect all information about personality traits of user activity within Facebook profiles using Facebook API. Based on all collected data, using the different classification technique they recognize the personality of user by their profile on Facebook. The data analysis was done by preprocessing the collected data. They compare all the results of all technique by calculating the F-measures of combining two indexes of precision and recall. The classifier of boosting-decision tree model gives the best accuracy with 82.2% than the previous techniques.

E. XGboosting

Michael M. et al [23] predicted the personality of user based on their social activities. They use the structure of social network and linguistic features relative to personality interactions using the mypersonality dataset. They analyzed and compared the different four machine learning models and perform the correlation between each of the feature set and personality traits. The personality prediction system built on the XGboosting classifier. They predict that XGboosting gives the more accuracy than the other classifier of the machine learning model. In comparison with the implementation of other gradient boosting, XGBoost avoids overfitting, and improves model performance and extraction speed.

V. Ong et al. [24] Predicted the personality based on Twitter in Bahasa Indonesia. Proposed model uses the two method of machine learning such as support vector machine and XGboosting. They use the 329 user's data to predict the personality. Experimental result using 10-fold cross validation shows that the system of using Xgboosting gives more accuracy than the Support vector machine.

Table -2: Overview of the personality prediction models and their algorithm

Dataset	Feature extraction	Feature selection	Methods	Ref
Facebook	psycho-linguistic LIWC	Chi-square	SVM	[18,22]
	LIWC	Pearson correlation	visualization	[25]
	LIWC, SNA, Time related post	-----	SVM+NB	[13]
	LIWC+CNN	Pearson correlation	CNN	[26]
	LIWC, SNA, SPLIC	Person correlation	XGboosting	[23,24]
	LIWC	-----	ZeroR	[22]
	linguistic information	-----	Multi-Label Semi-Supervised Classification	[13]
Text	psycho-linguistic		Supervised classifier	[18]
	Emotion related post	TF-IDF method	KNN	[14,15]
Bloggers	LIWC	-----	SVM	[17,18]
	n-gram	-----	Naïve Bayes	[27]

3. CONCLUSION

Social network analysis has increased largely in recent times. To extract the personality of the any person on the social networking websites is very useful for much application in various domain like including job success, attractiveness, and happiness. Personality detection from text means to extract the behavior characteristics of authors from the written text. This paper represents outline of insights for research on social networks and personality psychology. The study gives the literature about the various predicting algorithms which used to predict the personality of user from the social media

ACKNOWLEDGMENT

I Would like to thankful to Dept. of Computer Engineering GES's R.H. Sapat College of Engineering and Management Research.

REFERENCES

1. B.Y. Pratama and R Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM." In 2015 International Conference on Data and Software Engineering (ICoDSE), pp. 170-174. IEEE, 2015.
2. D. Sewwandi, K Perera, S Sandaruwan, O Lakchani, A Nugaliyadde and S Thelijjagoda, "Linguistic features-based personality recognition using social media data." In 2017 6th National Conference on Technology and Management (NCTM), pp. 63-68. IEEE, 2017.
3. M.M Tadesse, H Lin, B Xu and L Yang, "Personality Predictions Based on User Behavior on the Facebook Social Media Platform." IEEE Access 6 (2018): 61959-61969.
4. D Nie, Z Guan, B Hao, S Bai and T 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.
5. V Ong, A.D Rahmanto, D Williem and Suhartono, "Exploring Personality Prediction from Text on Social Media: A Literature Review." INTERNETWORKING INDONESIA 9, no. 1 (2017): 65-70.
6. E. Tupes and R. Christal, "Recurrent personality factors based on trait ratings," Journal of Personality, vol. 60, no. 2, pp. 225-251, 1992.
7. R. McCrae, "Why I advocate the five-factor model: Joint factor analyses of the NEO-PI with other instruments," Personality psychology: Recent trends and emerging directions, pp. 237-245, 1989.
8. J. Digman, "Personality structure: Emergence of the five-factor model Annual review of psychology," vol. 41, no. 1, pp. 417-440, 1990.
9. Alireza Souri, Shafiqeh Hosseinpour and Amir Masoud Rahmani, "Personality classification based on profiles of social networks' users and the five-factor model of personality" Human-centric computing and information Sciences springer, 2018.

10. Wei Z, Zhang H, Zhang Z, Li W, Miao D. "A naive Bayesian multi-label classification algorithm with application to visualize text Search Results" *Int J Advanced Intelligence*. 2011;3:173–88.
11. N.M.A Lestari, I.K.G.D .Putra, A.A.K.A. Cahyawan, "Personality types classification for Indonesian text I partners searching website using Naïve Bayes Methods" *International Journal of software and Informatics Issue*, 2013.
12. Lima, Ana CES, and Leandro N. De Castro. "Multi-label Semi-supervised Classification Applied to Personality Prediction in Tweets." *Computational Intelligence and 11th Brazilian Congress on Computational Intelligence (BRICS-CCI & CBIC)*, 2013 BRICS Congress on. IEEE, 2013.
13. G Farnadi, S Zoghbi, M.F Moens and M De Cock, "Recognizing personality traits using Facebook status updates." In *Seventh International AAI Conference on Weblogs and Social Media*. 2013.
14. Arifin, K. E. Purnaha, "classification of emotion in Indonesia text using KNN method" *International journal of Information and Electronics Engineering* 2.no 6, 2012.
15. Ryan T, Xenos S (2011) "Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage". *Comput Hum Behave* , 2011.27(5):1658–1664
16. Quercin D, Kosinski M, Stillwell D, Crowcroft J (2011)" Our twitter profiles, our selves: predicting personality with twitter". Paper presented at the privacy, security, risk and trust (PASSAT) and 2011 IEEE third international conference on social computing (SocialCom).
17. Iacobelli, F., Gill, A.J., Nowson, S. Oberlander, J. "Large scale personality classification of bloggers". In *Lecture Notes in Computer Science* (6975). 2011.
18. Soujanya Poria, Alexander Gelbukh, Basant Agarwal, Erik Cambria, Newton Howard, "Common Sense Knowledge Based Personality Recognition from Text", In *12th Mexican International Conference on Artificial Intelligence*, Volume 8266, 2013, pp 484- 496, Springer.
19. Y. Liu, J. Wang, and Y. Jiang, "PT-LDA: A Latent Variable Model to Predict Personality Traits of Social Network Users", *Neurocomputing*, vol. 210, pp. 155–163, 2016.
20. K. Peng, L. Liou, C. Chang and D. Lee, "Predicting personality traits of Chinese users based on Facebook wall posts," *2015 24th Wireless and Optical Communication Conference (WOCC)*, Taipei, 2015, pp. 9-14, doi: 10.1109/WOCC.2015.7346106.
21. Alireza Souri, Shafiqeh Hosseinpour and Amir Masoud Rahmani, "Personality classification based on profiles of social networks' users and the five-factor model of personality" *Human-centric computing and information Sciences* springer, 2018.
22. T . Tandra, D. Suhartono, R. Wongso, Y. L. Prasetyo et al., "Personality prediction system from facebook users," *Procedia Computer Science*, vol. 116, pp. 604–611, 2017.
23. M.M Tadesse, H Lin, B Xu, and L Yang, "Personality Predictions Based on User Behavior on the Facebook Social Media Platform." *IEEE Access* 6 (2018): 61959-61969.
24. S. Adali and J. Golbeck, "Predicting personality with social behavior," in *Advances in Social Networks Analysis and Mining (ASONAM)*, 2012 IEEE/ACM International Conference on. IEEE, 2012, pp. 302–309.
25. H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman et al., "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PloS one*, vol. 8, no. 9, p. e73791, 2013.
26. C Yuan, J Wu, H Li, and L Wang, "Personality Recognition Based on User Generated Content." In *2018 15th International Conference on Service Systems and Service Management (ICSSSM)*, pp. 1-6. IEEE, 2018.
27. Oberlander, J., and Nowson, S. 2006. "Whose thumb is it anyway? classifying author personality from weblog text". In *Proc. of the 44th Annual Meeting of the Association for Computational Linguistics ACL*. 627–634 .