



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 6, June 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.512

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Personality Prediction via Social Media Using Data Mining

Miss.Jayita Bhattacharjee, Miss. Chandrakala Kale, Miss. Dipalirakhunde, Miss. Priyanka More,
Prof.Pankaj Khambre

G.H.Raisoni Institute of Engineering and Technology, Pune, Maharashtra, India

ABSTRACT: with the development of social networks, a large variety of approaches have been developed to define users' personalities based on their social activities and language use habits. Particular approaches differ with regard to different machine learning algorithms, data sources and feature sets. The goal of this paper is to investigate the predictability of the personality traits of social media users based on different features and measures. We extract social data and questionnaire, and focus on how to use the user text information to predict their personality characteristics. We use the correlation analysis and principal component analysis to select the user information, and then use the multiple regression model, the gray prediction model and the multitasking model to predict and analyze the results.

KEYWORDS: Data Mining, Personality Prediction, sentiment analysis, social media.

I. INTRODUCTION

The popularity of social networks makes people's social changes, making the exchange of people, communication and cooperation between the changes. On the one hand, people can use the network platform to contact friends, comment, and discuss public topics and so on. On the other hand, the role of social networks in daily life is increasing, and even to a large extent affected the reconstruction of the network of real social relations. At the same time, because the behavior and status of social networks are easily recorded, acquired and analyzed, social computing has become an important research content in the field of information technology and computer. Because people's behavior and personality are closely linked, making personality prediction has a broad academic value and business prospects. Thus, the use of social networking to tap the user's social information has become the focus of attention of enterprises and scholars. Personality is a high degree of generalization of different individual characteristics of mankind, even in the same environment, different people will show different behavior, which comes from the different personality of each person. Personality psychology is one of the branches of psychology, mainly through people's external behavior to distinguish between people's intrinsic characteristics, and to study the relationship between them. Psychology usually use personality traits to define people's personality, explain the user's behavior and preferences.

The organization of the paper is as follows section II gives the related work and limitations and last section concludes the paper with future work followed by references.

II. RELATED WORK

There is a growing number of research papers related to a user's behaviour in social networks that has recently attracted more attention in the international research community. Personality recognition is studied by two main disciplines: computational linguistics and Social Network Analysis. From the area of computational linguistics.

Pennebaker and King 1999 [2] wrote a pioneering work dedicated to personality extraction from text. They examined words in a variety of domains such as diaries, college writing assignments and social psychology manuscripts to study personality related features with linguistic cues. Their results show that agreeable people tend to use more articles while introverts and those low in conscientiousness use more words signalling distinctions. Neurotics use more negative emotion words.



Argamon et al. 2005 [3] classified neuroticism and extraversion using linguistic features such as function words, judgemental and appraisal expressions and modal verbs. Their results revealed that neuroticism is related to the use of functional lexical features, for instance appraisal lexical taxonomy, whereas the results for extraversion were less clear.

Other studies linked neuroticism to irrational beliefs or poor coping efforts on well-being personality [4] oberlander and Nowson 2006 classified the extraversion, stability, agreeableness and conscientiousness of bloggers using the Naive Bayes prediction model as a learning algorithm using different sets of n-grams as features.

Mairesse et al. 2007 [5] examined correlations between the Big 5 personality traits, using LIWC and RMC as feature sets. While LIWC features included word classification such as positive emotions or anger, RMC features included results about word age of acquisition or word imageability.

In Social Network Analysis, personality recognition extracted from network configuration and other extra-linguistic cues has an even shorter history. The impact of a user’s social interaction behaviour on personality was studied by Gosling et al. [9].

III. PROPOSED SYSTEM

We extract social data and questionnaire, and focus on how to use the user text information to predict their personality characteristics. We use the correlation analysis and principal component analysis to select the user information, and then use the multiple regression model, the prediction model and the multitasking model to predict and analyse the results.

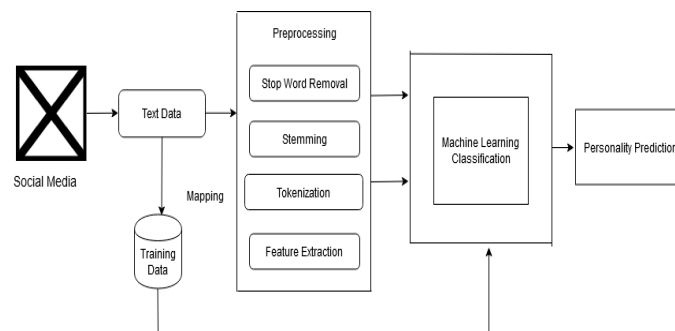


Fig. System Architecture

Algorithm

Naive Bayes

Steps:

1. Given training dataset D which consists of documents belonging to different class say Class A and Class B
2. Calculate the prior probability of class A=number of objects of class A/total number of objects
Calculate the prior probability of class B=number of objects of class B/total number of objects
3. Find NI, the total no of frequency of each class
Na=the total no of frequency of class A
Nb=the total no of frequency of class B
4. Find conditional probability of keyword occurrence given a class:
 $P(\text{value1}/\text{ClassA}) = \text{count}/\text{ni}(A)$
 $P(\text{value1}/\text{ClassB}) = \text{count}/\text{ni}(B)$
 $P(\text{value2}/\text{ClassA}) = \text{count}/\text{ni}(A)$



$$P(\text{value2/ClassB}) = \text{count}/n_i(B)$$

.....
.....
.....

$$P(\text{valuen/ClassB}) = \text{count}/n_i(B)$$

5. Avoid zero frequency problems by applying uniform distribution
6. Classify Document C based on the probability $p(C/W)$
 - a. Find $P(A/W) = P(A) * P(\text{value 1/Class A}) * P(\text{value 2/Class A}) * \dots * P(\text{value n /Class A})$
 - b. Find $P(B/W) = P(B) * P(\text{value 1/Class B}) * P(\text{value 2/Class B}) * \dots * P(\text{value n /Class B})$
7. Assign document to class that has higher probability.

IV. RESULTS AND ANALYSIS

The Experiments are done by personal computer with configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and JDK 1.8. The application uses web application tool for design code in Eclipse and execute on Tomcat server.

Positive (P) : Observation is positive.

Negative (N) : Observation is not positive.

True Positive (TP) : Observation is positive, and is predicted to be positive.

False Negative (FN) : Observation is positive, but is predicted negative.

True Negative (TN) : Observation is negative, and is predicted to be negative.

False Positive (FP) : Observation is negative, but is predicted positive.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

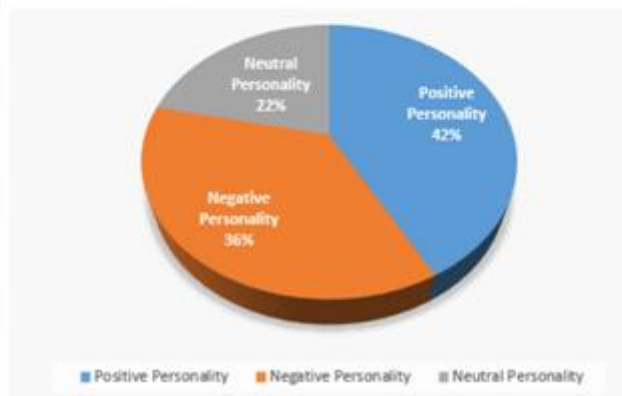
$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1\text{-Measure} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall}).$$



	Naive Bayes(NB)	Support Vector Machine(SVM)
Precision	68.45	78.70
Recall	79.44	65.64
F-Measure	72.11	74.31
Accuracy	80.29	87.26



Types of personalities	Results
Positive Personality	42%
Negative Personality	36%
Neutral Personality	22%

V. CONCLUSION

In this paper, we provide an outline of insights for research on social networks and personality psychology. The study investigates the literature on the uses of social media frameworks as behavioural feature study by exploring the relationship between users’ personalities and their behaviours in social networks. To predict a user’s personality, we conducted a comparative study of best behavioural indicators for Facebook usage of the same set of features to capture the ways the users socialize, communicate and connect with each other.

REFERENCES

- [1] M. S. H. Mukta, M. E. Ali, and J. Mahmud, “User generated vs. Supported contents: Which one can better predict basic human values?” in Proc. Int. Conf. Social Inform. Cham, Switzerland: Springer, 2016, pp. 454–470.
- [2] P. Howlader, K. K. Pal, A. Cuzzocrea, and S. D. M. Kumar, “Predicting Facebook-users’ personality based on status and linguistic features via flexible regression analysis techniques,” in Proc. 33rd Annu. ACM Symp. Appl. Comput. (SAC), Pau, France, Apr. 2018, pp. 339–345.
- [3] KAUSHAL AND M. PATWARDHAN, “EMERGING TRENDS IN PERSONALITY IDENTIFICATION USING ONLINE SOCIAL NETWORKS—A LITERATURE SURVEY,” ACM TRANS. KNOWL. DISCOV. DATA, VOL. 12, NO. 2, PP. 1–30, JAN. 2018.
- [4] W. Youyou, M. Kosinski, and D. Stillwell, “Computer-based personality judgments are more accurate than those made by humans,” Proc. Nat. Acad. Sci. USA, vol. 112, no. 4, pp. 1036–1040, Jan. 2015.
- [5] G. Farnadi, G. Sitaraman, S. Sushmita, F. Celli, M. Kosinski, D. Stillwell, S. Davalos, M.-F. Moens, and M. De Cock, “Computational personality recognition in social media,” User Model. User-Adapted Interact., vol. 26, nos. 2–3, pp. 109–142, 2016.
- [6] S. Mallik, T. Bhadra, and U. Maulik, “Identifying epigenetic biomarkers using maximal relevance and minimal redundancy based feature selection for multi-omics data,” IEEE Trans. Nanobiosci., vol. 16, no. 1, pp. 3–10, Jan. 2017, doi: 10.1109/tnb.2017.2650217.



- [7] T. Tandra, D. Suhartono, R. Wongso, Y. L. Prasetio et al., "Personality prediction system from facebook users," *Procedia Computer Science*, vol. 116, pp. 604–611, 2017.
- [8] ONG, A. D. RAHMANTO, D. SUHARTONO, A. E. NUGROHO, E. W. ANDANGSARI, M. N. SUPRAYOGI ET AL., "PERSONALITY PREDICTION BASED ON TWITTER INFORMATION IN BAHASAINONESIA," IN *COMPUTER SCIENCE AND INFORMATION SYSTEMS (FEDCSIS)*, 2017 FEDERATED CONFERENCE ON. IEEE, 2017, pp. 367–372.
- [9] F. Iatan, "Predicting human personality from social media using a fuzzy neural network," in *Issues in the Use of Neural Networks in Information Retrieval*. Springer, 2017, pp. 81–105
- [10] T. Bhadra, S. Mallik, and S. Bandyopadhyay, "Identification of multiview gene modules using mutual information-based hypograph mining," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 6, pp. 1119–1130, Jun. 2019, doi: 10.1109/tsmc.2017.2726553.



INNO SPACE
SJIF Scientific Journal Impact Factor

**Impact Factor:
7.512**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details