



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

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 9, Issue 12, December 2020

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.512**

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Personality Prediction from Real Time Tweets using NLP

Vijayalaxmi Tadkal<sup>1</sup>, Suman Patil<sup>2</sup>, Sangeeta Thakur<sup>3</sup>, Kaveri N M<sup>4</sup>, Neeta P Jidage<sup>5</sup>

Asst. professor, Department of Information Science and Engineering, Godutai Engineering College for Women,  
Kalaburagi, Karnataka, India<sup>12</sup>

B.E Students, Department of Information Science and Engineering, Godutai Engineering College for Women,  
Kalaburagi, Karnataka, India<sup>345</sup>

**ABSTRACT:** Digital networks has caught the interest of individuals of the present generation and has become very available on the internet. The activities carried out by users of social media provide researchers with a greater forum to study and learn about online habits, interests, and characteristics. Since the number of users is high, we use Twitter as a leading social media site for research. User data can be collected and performed in order to evaluate the essence of the individual Further analysis is based on the email, status profiles, or preferences provided on Twitter to other users. Comments added by the user, tweets, social actions, and the usage of language patterns on Twitter are taken as data for review in order to understand the various types of text. Using Machine Learning algorithms to process the data and the real predictive personality of the user will take this research further.

**KEYWORDS:** Network, Digital, Tweets, Machine Learning.

## I. INTRODUCTION

Platforms for social networks play an important role in gaining the interest of people who are interested on the Internet. A group of researchers have found that many individuals are involved on the social media site to explore the various personalities of different individuals in their everyday lives. Using comments, tweets, shared contacts, and any form of text messaging, this information can be collected. By using these kinds of processed data, other profiles may denote the online activity of the user. Our research initially focuses on social behavior and twitter language patterns. For each dimension of personality, we select the most useful characteristics and predict the character of the user. Centered on two categories, we suggested a systematic method. One is the study of the social network(SNA) and two classes of linguistic inquiry and word count(LIWC) and organized programming based on the dataset we used for linguistic cue extraction(SPLICE). By processing the data, we can check some of the similarities shown based on these categories. The correlations between personality traits and sets of characteristics are almost the same. It is easier to predict some simple norms. We came to a conclusion to process the data for information by using multiple sets of categories such as network size, tweet density, occupation, and number of connections. The predictive existence of the characteristics set by predictions using personality traits has been explored. We have used a parallel Machine Learning algorithm to continue with the implementation model to traverse the degree to which we can predict the Twitter personality traits

## II. LITERATURE SURVEY

Pennebaker et al. [2] explained the work of personality extraction from the text. They have collected the data through a manual process such as dairy writings, college assignments, and psychology manuscripts to study the personality of a person. Their outcomes are little satisfied for extravert person and not for introverts. Usually, introverts don't show much interest in these practices. So, this fails in the broader area.

Argamon et al. [3] worked on extraversion and neuroticism using linguistic features using some categories of traits. They are appraisal, function words, expressions, modal verbs, etc. In this work also, failure is caused due to some lesser outcomes. It lags in the result of extraversion and mainly used for functional lexical features. Oberlander and Nowson [4] classified the personality traits using the style of language habits from bloggers using a Machine learning algorithm called Naïve Bayes to predict it. This has been applied using different sets of n-gram as features, and it is feeble in the copying of efforts on wellbeing personality[6].

Kalish and Robins [5][7] examined the personalities of individuals by working on their immediate network. They both were working on the egoist character of the environment. By focusing on this, they were successfully collected the datasets of more person's character. Their analysis concluded that psychological predispositions could be explained by the dataset even though the text or language-based data has received much attention and more efficiency through other research.

### III. PROPOSED WORK

In proposed system, the dataset obtained from Twitter social media platform. These data are preprocessed and passed into Feature Extraction and then Feature Selection. And finally, the extracted features are trained by the XGboost algorithm, and the resultant classifies the personality with the accuracy.

### IV. METHODOLOGY

#### Dataset And Dataset Preprocessing

From the twitter API datasets are collected. This helps us to make use of Twitter functionality without the aid of the website gui. The Twitter API can be used in an automated way to post tweets or to send messages. The API is also regarded as a list of URL documentation. Through this, we have some knowledge to predict the personality of the people.

Before it advanced to the feature selection and training level, the dataset obtained from my personality was pre-processed. We used OpenNLP to pre-process the dataset. First, we used tokenization, including punctuation and aggregation of similar words, to separate the last word of each sentence. Next, URLs, icons, names, spaces and lower cases have been deleted.

Before it advanced to the feature selection and training level, the dataset gathered from my personality was pre-processed. We used OpenNLP in order to pre-process the dataset. First, we used tokenization, including punctuation and aggregation of similar words, to break the last word of each sentence. Next, URLs, icons, names, spaces, and lower cases have been deleted.

The behavior of individuals on social media networks is correlatively influenced by the behavior and involvement of other individuals on the web. This can have an effect on the creation of new data or behavior by the Yeah. Groups.

#### Feature Extraction

The behavior of individuals on social media networks is correlatively influenced by the behavior and involvement of other individuals on the web. This will generate an impact through the groups on the creation of new data or behavior. Here classifying the dataset into two groups, the first one is text feature extraction which is employed to analyse the people's language on Twitter it also includes two features namely topic and expression counts.

The LIWC(Linguistic Inquiry and Word Count) is potentially used to understand the psychology of the people, this tool is vast spread for psychological studies. LIWC2015 is mainly developed to analyse the files in many languages in a timely period and efficiently[9]. SPLICE(Structured Programming for Linguistic Cue Extraction) it is a analysis tool for linguistic which is under testing stage.

It can be used for personality prediction when it is completely developed. Feature extraction is done by using NLTK(Natural Language ToolKit).NLTK is an nlp library that comprises packages to do understand human communication to the machine. Package includes stemming, lemmatization, tokenization, POS tagging, extraction[10].

#### Feature Selection

Usually, there are a few primary reasons why the choice of features is important for the creation of a model[11]. It reduces the high dimensionality of the dataset by removing the features that are not required for training, reinforcing the generalization of the model, and thus reducing the training time. Then the model acquires a sound understanding of the features and their relationships to the features of the response. Added to this, the performance of the training algorithms is also enhanced and the processing requirements are reduced[12]. For evaluating the linear dependence between two quantitative or continuous variables, we have used the pearson correlation method. It is also known as the statistical test for Pearson R. The value returned is between -1 and +1 here.

## V. CONCLUSIONS

In this report, we have provided a structure for insights into the social network analysis and personality prediction of each consumer. Using the user's viewpoint on the relationship, tweets, and feedback on the social network, our survey on this routine study operation is effectively completed. In order to predict the essence or character of a user, we have supervised a comparative analysis of the best behavioral standards for the use of the same set of features on Twitter to document how the user socializes, interacts and connects. We have used a large range of datasets to evaluate the various personalities of an individual for a successful outcome. Our findings show that studying the social and linguistic network of personality will gain a more significant number of insights.

## REFERENCES

- [1] J. Staiano, B. Lepri, N. Aharony, F. Pianesi, N. Sebe, and A. Pentland, "Friends don't lie: Inferring personality traits from social network structure," in Proc. ACM Conf. Ubiquitous Comput., 2012, pp. 321–330.
- [2] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," Tech. Rep., 2015.
- [3] T. P. Michalak, T. Rahwan, and M. Wooldridge, "Strategic social network analysis," in Proc. AAAI, 2017, pp. 4841–4845.
- [4] S. Argamon, S. Dhawle, M. Koppel, and J. Pennebaker, "Lexical predictors of personality type," Tech. Rep., 2005.
- [5] Y. Kalish and G. Robins, "Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure," Social Netw., vol. 28, no. 1, pp. 56–84, 2006.
- [6] R. R. McCrae and P. T. Costa, "Personality, coping, and coping effectiveness in an adult sample," J. Pers., vol. 54, no. 2, pp. 385–404, 1986.
- [7] P. T. Costa and R. R. McCrae, "Normal personality assessment in clinical practice: The NEO personality inventory," Psychol. Assessment, vol. 4, no. 1, pp. 5–13, 1992.
- [8] S. Adali and J. Golbeck, "Predicting personality with social behavior," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2012, pp. 302–309.
- [9] D. J. Brass, "Being in the right place: A structural analysis of individual influence in an organization," Admin. Sci. Quart., vol. 29, no. 4, pp. 518–539, 1984.
- [10] R. Wald, T. M. Khoshgoftaar, A. Napolitano, and C. Sumner, "Using Twitter content to predict psychopathy," in Proc. 11th Int. Conf. Mach. Learn. Appl. (ICMLA), vol. 2, Dec. 2012, pp. 394–401.
- [11] M. A. Hall, Correlation-Based Feature Selection for Machine Learning. 1999.
- [12] G. Farnadi et al., "How are you doing?: Emotions and personality in Facebook," in Proc. EMPIRE (2nd Workshop Emotions Personality Personalized Services), Workshop UMAP (22nd Conf. User Modeling, Adaptation Personalization), 2014, pp. 45–56.
- [13] S. Bai, R. Gao, and T. Zhu, "Determining personality traits from Renren status usage behavior," in Computational Visual Media. Springer, 2012, pp. 226–233.
- [14] H. Wei et al., "Beyond the words: Predicting user personality from heterogeneous information," in Proc. 10th ACM Int. Conf. Web Search Data Mining, 2017, pp. 305–314.
- [15] Y. Amichai-Hamburger and G. Vinitzky, "Social network use and personality," Comput. Hum. Behav., vol. 26, no. 6, pp. 1289–1295, 2010.



**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](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details