PERSONALITY PREDICTION SYSTEM USING MACHINE LEARNING

Ms. C. Ashwini, Sk Shahid Ali, Arnab Rooj
Assistant Professor, Pre-final Year Student, Pre Final Year Student
Computer Science and Engineering Department
SRM Institute Of Science And Technology, Ramapuram, Chennai, India.

Abstract — A wide range of methods have been created with the growth of social networks to determine the personalities of consumers based on their social activities and practices of language use. In terms of distinct machine learning algorithms, information sources, and function sets, particular methods vary. The purpose of this article is to explore the predictability of Facebook users ' personality characteristics. The actions of these users on social networks provide scientists with a excellent platform to study and comprehend their behaviors, preferences and personalities online. Our research predicts personality based on the social behavior of consumers on Facebook's social media platform and their language-use practices. With the greatest correlations, we examine the characteristics.

Index Terms — Machine Learning, Feature analysis, predicting personality, social behavior, social networks.

I. INTRODUCTION

Over the previous few years, social media has become the most commonly used instrument for communication and interaction among individuals. Direct people-to-people interaction is declining as people communicate indirectly via smartphones. Thus, recognizing the personality of the person is quite difficult. What's written in social media, however, could help us get the information needed as people spend a lot of time checking social media and expressing their feelings and thoughts through status, comments, and updates. Facebook has the largest users reaching 1.8 billion users, with approximately 800 million users spending about 40 minutes a day 1. Several personality models are used for personality prediction, such as Big Five Personality, MBTI (MyersBriggs Type Indicator) or DISC (Dominance Influence Steadiness Consciousness). However, the most prominent and promising method after some considerations and literature review process is the use of Artificial Neural Networks, which is essentially an application for machine learning. Machine learning includes artificial intelligence that empowers the system to learn and enhance from past experiences without ever programming. The characteristics of this model are Openness, Consciousness, Extraversion, Agreeability, and Neuroticism.

Our study has some specific contributions: first, to clarify the relationship between user personalities and their behavior in social networks; second, to illustrate a higher potential for personality prediction of individual social network features by using Logistic Regression and KNN. An artificial or convolutionary neural network is a class of deep neural networks in machine learning, most commonly used for visual imaging analysis. In machine learning , logistic regression models the probabilities for classification problems with two possible outcomes. It's an extension of the linear regression model for classification problems. Logistic regression is a classification algorithm used to assign observations to a discrete set of classes. Unlike linear regression which outputs continuous number values, logistic regression transforms its output using the logistic sigmoid function to return a probability value which can then be mapped to two or more discrete classes. The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems. The KNN algorithm assumes that similar things exist in close proximity.

II. PROBLEM DESCRIPTION

Over the past few years, social media has become the most commonly used tool for communication and interaction between people. Direct people-to-people contact is growing as people communicate indirectly through smart phones. Therefore, identifying the personality of the individual is quite difficult. What's written in social media, though, can help us get the information needed because people spend a lot of time monitoring social media and sharing their feelings and thoughts through status, tweets, and notifications. The purpose of this research is to build a system of predictions that can automatically predict user personality based on their Facebook activities. Many personality models are used for personality forecasting, such as Big Five Personality, MBTI (MyersBriggs Type Indicator) and DISC (Dominance Impact Steadiness Consciousness). Nonetheless, Big Five Personality is used in this study after some considerations and literature review process as it is the most common and effective in telling somebody's personality traits. The characteristics of this model are Openness, Consciousness, Extraversion, Agree-ability, and Neuroticism.

III. LITERATURE SURVEY

1. Friends don't Lie - Inferring personality traits from social network structure

In this job, we explore the interactions between the structure of social networks and personality; in anticipating theBig-5 personality traits, we evaluate the performance of distinct subsets of organizational network characteristics, especially those involved with ego-networks. In addition to traditional survey-based data, this work focuses on social networks derived from real-life data gathered through smartphones. Besides showing that the latter are superior to the former for the task at hand, our results provide a fine-grained analysis of the contribution the various feature sets are able to provide to personality classification, along with an assessment of the relative merits of the various networks exploited.

2. The development and psychometric properties of LIWC

In their daily life, the ways individuals use words can provide rich data about their convictions, fears, patterns of thinking, social interactions, and personalities. From the moment of Freud's texts about tongue slips to the early days of computer-based text assessment, scientists started to gather ever more compelling evidence of tremendous psychological significance from the phrases we use. In order to provide an efficient and effective method for studying the various emotional, cognitive, and structural components present in individuals' verbal and written speech samples, we originally developed a text analysis application called Linguistic Inquiry and Word Count, or LIWC.

3. Lexical Predictors of personality types

We are presently pursuing "author profiling" techniques in which different aspects of the author's identity can be identified from a document without necessarily getting an acorpus of the same individual's records. Personality is a main element of such an identity profile; this article discusses the distinction between elevated and low neuroticism and extraversion in casual text writers. For this assignment, we consider four distinct sets of lexical features: a normal word list function, conjunctive sentences, indicative modality, and adjectives and modifiers for evaluation. SMO, a vector machine learner supported, was used in each of the two tasks to learn linear separators for the high and low classes. We find that the use of evaluation is the greatest predictor of neuroticism, and that word feature works best for extraversion. In addition, examining the specifically most significant fea-tures provides insight into how neuroticism and extraversion impact language use differently.

4. Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information

Despite the huge popularity of Online Social Networking sites (OSNs; e.g., Facebook and Myspace), little psychological study has been performed on them. Two studies examining how personality is reflected in OSNs disclosed several associations between the characteristics of the Big Five and self-reported Facebook behaviors and observable profile data. For example, extraversion predicted not only frequency of Facebook usage, but also engagement in the site, with extraverts (vs. introverts) showing traces of higher levels of Facebook activity. As in off-line situations, extraverts seek virtual social engagement, leaving behind a residue of behavior in the form of lists of friends and posts of pictures. Results indicate that OSN users appear to extend their off-line personalities to OSN domains rather than fleeing or compensating for their off-line character.

5. The Five-Factor Model of personality and Degree and Transitivity of Facebook social networks.

This research examined the connections of the personality aspects of the Five-Factor Model with Degree (number of colleagues) and Transitivity (like two of aperson's colleagues being friends with each other) of a person's network on Facebook's social networking site. A Facebook request administered a questionnaire on personality and objectively recorded social networks of the respondents (N= 5031;65% female, mean age 33 years). Extraversion predicted degree as anticipated, while Openness to Experience predicted more cross-sex friendships and Agreeableness among males. Also aspected, Extraversion and, among males, Openness to Experience predicted reduced Transitivity, but when Degree was controlled for, the former association was made insignificant.

IV. DISADVANTAGES OF THE EXISTING SYSTEM

- Decision boundaries may be over-trained, which means that if our training set does not include some examples that we want to have in a class, after training we may not get the correct class label.
- The accuracy of the prediction method is comparatively low.
- Computation needs a lot of time, so do the classification.

.

V. PROPOSED SYSTEM

In machine learning, logistic regression models with two possible outcomes the probability of classification problems. Logistic regression is used to classify data and explain the relationship between one dependent binary variable and one or more independent nominal, ordinal, interval or ratio type variables. Often logistic regressions are difficult to interpret; you can easily evaluate the Intellectus Statistics method and then interpret the results in plain English. The task of estimating the log odds of an event is at the center of the logistic regression analysis. Mathematically, a multiple linear regression function is estimated to be defined as:

$$= \log\left(\frac{p(y=1)}{1-(p=1)}\right) = \beta_0 + \beta_1 x_{i2} + \beta_2 \cdot x_{i2} + \_ + \beta_p \cdot x_{in}$$

The K-nearest Neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm which can be used to solve problems of classification and regression. The KNN algorithm assumes that in close proximity there are similar things. KNN algorithm hinges on the assumption that the algorithm is sufficiently true to be useful. KNN captures the concept of similarity (sometimes referred to as distance, proximity or proximity) with some of the mathematics we may have learned in our childhood, calculating the distance between points on a map.

VI. METHODOLOGIES

1. KNN(THE K NEAREST NEIGHBOURS)

The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems. The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other. The KNN algorithm hinges on this assumption being true enough for the algorithm to be useful. We are given some prior data (also called training data), which classifies coordinates into groups identified by an attribute. KNN captures the idea of similarity (sometimes called distance, proximity, or closeness) with some mathematics we might have learned in our childhood — calculating the distance between points on a graph. The KNN Algorithm:

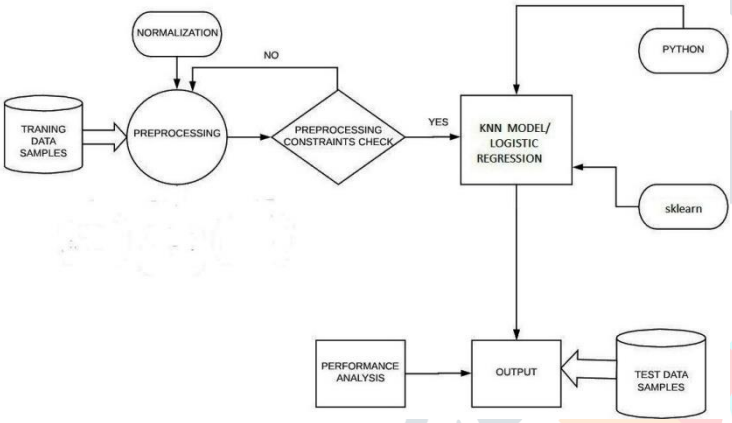
1. Load the data
2. Initialize K to your chosen number of neighbors
3. For each example in the data
Calculate the distance between the query example and the current example from the data.
Add the distance and the index of the example to an ordered collection
4. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances
5. Pick the first K entries from the sorted collection
6. Get the labels of the selected K entries
7. If regression, return the mean of the K labels
8. If classification, return the mode of the K labels

2 . LOGISTIC REGRESSION

Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. Sometimes logistic regressions are difficult to interpret; the Intellectus Statistics tool easily allows you to conduct the analysis, then in plain English interprets the output.

VII. SYSTEM ARCHITECTURE

First step : Extraction of feature is the method of reducing the original set of enormous raw information to more manageable processing groups. Large set of information involve the processing of a lot of computing resources. Second step : Classification determines which category belongs to the information set. Training information for training an algorithm is used. Test information are used to evaluate model output and enhance precision. Third step : To predict the personality,we use qualified information set results.



VIII. MODULE IDENTIFICATION

DataCollection:
The process of collecting and measuring information from Facebook is data collection. It must be collected and stored in a way that makes sense to use the information we gather to create realistic artificial intelligence and machine learning solutions. Our data consists mainly of the Facebook user dataset labeled in accordance with the Big 5 model.

Pre-processing:
Pre-processing refers to the transformations applied to our data before feeding it to the algorithm.
Data pre-processing is a part of data mining, which involves transforming raw data into a more coherent format.
Data Pre-processing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw

Training the Machine:
To suit the prototypes, the training sets are used. The model's training includes cross-validation where we use training data to get the model's approximate performance. The training process of an ML model involves providing training data to learn from an ML algorithm (i.e., the learning algorithm).

Data Scoring:
Scoring is also called prediction and, given some new input data, is the process of generating values based on a trained model of machine learning.
Logistic Regression And KNN is the technique used to process the dataset. The values or scores created can represent an individual's personality predictions.

IX. EXPERIMENT RESULTS

Logistic Regression And KNN is the technique used to process the data set. The values or scores that are created can represent an individual's personality predictions. By exploring the relationship between the personalities of users and their behaviors in social networks, the study investigates the literature on the uses of social media framework as a behavioral feature study. Our results show that it is possible to gain a great deal of insight from studying personality social and linguistic indicators. Developing these approaches and evaluating them is a new space open to future work. The following figures show the experiment's output:

	Gender	Age	openness	neuroticism	conscientiousness	agreeableness	extraversion	Personality (Class label)
2	Male	17	7	4	7	3	2	extraverted
3	Male	19	4	5	4	6	6	serious
4	Female	18	7	6	4	5	5	dependable
5	Female	22	5	6	7	4	3	extraverted
6	Female	19	7	4	6	5	4	lively
7	Male	18	5	7	7	6	4	lively
8	Female	17	5	6	5	7	4	extraverted
9	Female	19	6	6	7	5	4	extraverted
10	Male	18	5	7	5	6	7	dependable

Fig. 1

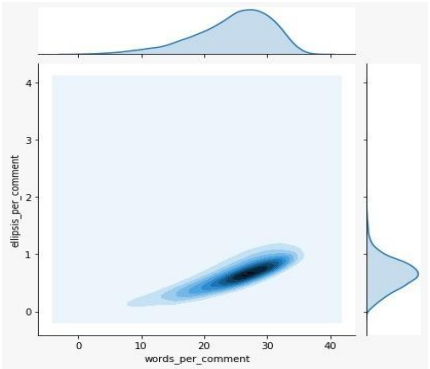


Fig. 2

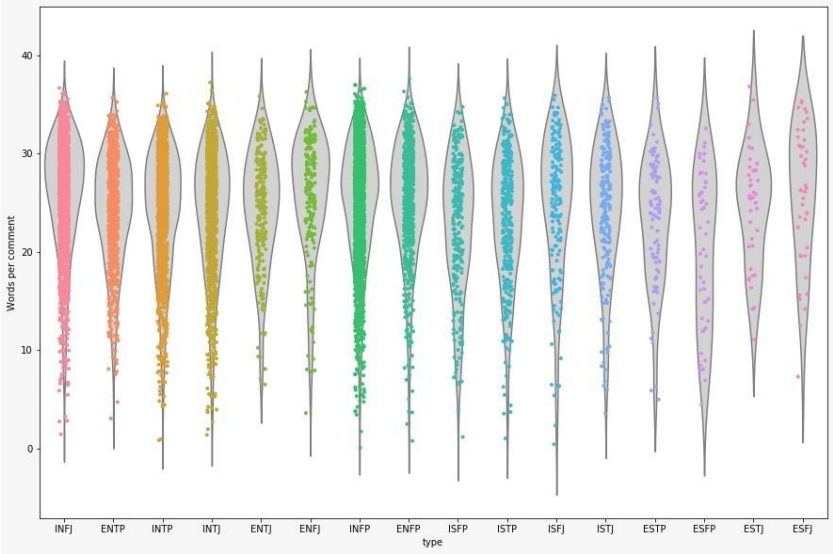


Fig. 3

X. 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. 321330.

[2] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, ``The development and psychometric properties of LIWC2015," Tech. Rep.,2015.

[3] S. Argamon, S. Dhawle, M.Koppel, and J. Pennebaker, ``Lexical predictors of personality type," Tech. Rep., 2005.

[4] I. 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. 81105.

[5] H. A. Schwartz 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.

[6] T. Morton, J. Kottmann, J. Baldridge, and G. Bierner, ``Opennlp: A javabased NLP toolkit," in Proc. EACL, 2005.

[7] V. Evrim and A. Awwal, ``Effect of personality traits on classication of political orientation," World Acad. Sci., Eng. Technol., Int. J. Social, Behav., Educ., Econ., Bus. Ind. Eng., vol. 9, no. 6, pp. 20012006, 2015.