PREDICTING PERSONALITY FROM TWITTER USING MACHINE LEARNING

Mrs. M. Sruthi [1]., G. Shreya[2]., K. Bhavani Shankar[3]., T. Sumanth Rayudu[4].

[1] Assistant Professor, [2,3,4]Students B. Tech -IT, (20S11A1237, 20S11A1210, 20S11A1242).

Malla Reddy Institute of Technology and Science., Maisammaguda., Medchal., Ts, India

[1]Sruthi.mukkala@mrits.ac.in, [2]shreya.gadhe2002@email.com, [3]shankarkaja0791@gmail.com, , [4]sumanthrayudu0@gmail.com.

ABSTRACT

Social media is a place where users present themselves to the world, revealing personal details and insights into their lives. We are beginning to understand how some of this information can be utilized to improve the users' experiences with interfaces and with one another. In this paper, we are interested in the personality of users. Personality has been shown to be relevant to many types of interactions; it has been shown to be useful in predicting job satisfaction, professional and romantic relationship success, and even preference for different interfaces. Until now, to accurately gauge users' personalities, they needed to take a personality test. This made it impractical to use personality analysis in many social media domains. In this paper, we present a method by which a user's personality can be accurately predicted through the publicly available information on their Twitter profile. We will describe the type of data collected, our methods of analysis, and the machine learning techniques that allow us to successfully predict personality. We then discuss the implications this has for social media design, interface design, and broader domains.

1. INTRODUCTION

Social networking on the web has grown dramatically over the last decade. In January 2005, a survey of social networking websites estimated that among all sites on the web there were roughly 115 million members [14]. Just over five years later, Twitter alone has exceeded 200 million members. In the process of creating social networking profiles, users reveal a lot about themselves both in what they share and how they say it. Through self-description, status updates, photos, and interests, much of a user's personality comes out through their profile. For decades, psychology researchers have worked to understand personality in a systematic way. After extensive work to develop and validate a widely accepted personality model, researchers have shown connections between general personality traits and many types of behavior. Relationships have been discovered between personality and psychological disorders, job performance and satisfaction, and even romantic success. This paper attempts to bridge the gap between social media and personality research by using the information people reveal in their online profiles. Our core research question asks whether social media profiles can predict personality traits. If so, then there is an opportunity to integrate the many results on the implications of personality factors and behavior into the users' online experiences and to use social media profiles as a source of information to better understand individuals. For example, the friend suggestion system could be tailored to a user based on whether they are more introverted or extraverted. Previous work has shown that the information in users' Facebook profiles is reflective of their actual personalities, not an "idealized" version of themselves. We expect Twitter to have similar characteristics, and that plus a broad user base of 200 million people makes it an ideal platform for study. We administered the Big Five Personality Inventory to 279 subjects through a Twitter application. In the process, we gathered their 2000 most recent public Twitter posts (tweets). This was aggregated, quantified, and passed through a text analysis tool to obtain a feature set. Using these statistics, we were able to develop a model that can predict personality on each of the five personality factors to within between 11% and 18% of the actual values. The ability to predict personality has implications in many areas. Existing research has shown connections between personality traits and success in both professional and personal relationships. Social media tools that seek to support these relationships could benefit from personality insights. Additionally, previous work on personality and interfaces showed that users are more receptive to and have greater trust in interfaces and information that is presented from the perspective of their own personality features (i.e., introverts prefer messages presented from an introvert's perspective). If a user's personality can be predicted from their social media profile, online marketing and applications can use this to personalize their message and its presentation. We begin by presenting background on the Big Five Personality index and related work on personality and social media. We then present our experimental setup and methods for analyzing and quantifying Twitter profile information. To understand the relationship between personality and social media profiles, we present results on correlations between each profile feature and personality factor. Based on this, we describe the machine learning techniques used for classification and show how we achieve large and significant improvements over baseline classification on each personality factor. We conclude with a discussion of the implications that this work has for social media websites and for organizations that may utilize social media to better understand the people with whom they interact.

2. LITERATURE SURVEY

1.Online Social Networks and Insights into Marketing Communications: Even though online social network services have become enormously popular among public,
there is a laxity of empirical investigations on the individual's level in this domain. This paper
examines the impact of personality factors such as extraversion, self-esteem, opinion seeking and opinion leadership on brand communication and online social behaviours. Our results show that gender and extroversion predict online social network size and time spent online; that opinion seekers spend more time online and have larger networks relative to opinion leaders and that opinion leaders are more likely to communicate their brand use online. We also find the mediating role of opinion leadership and opinion seeking in explaining the impact of general personality traits on online brand communication and social networking. Directions for future research are provided and some practical implications are discussed.

2. Facebook Profiles Reflect Actual Personality, Not Self Idealization: This study examined whether profiles in online social networking sites (OSNs) convey accurate impressions of profile owners. Participants were 236 OSN users from the most popular OSNs in the United States (Facebook, N = 133) and Germany (StudiVZ, SchuelerVZ; N = 103). In the U.S. sample, profile owners' self-reports and reports from four well-acquainted friends were obtained using the Ten Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003). In the German sample, self-reports on the short form of the Big Five Inventory (BFI-10; Rammstedt & John, 2007) and the NEO Five-Factor Inventory (Costa & McCrae, 1992) were combined. Our results were consistent with the extended real-life hypothesis and contrary to the idealized virtual-identity hypothesis. Observer accuracy was found, but there was no evidence of self-idealization (see Table 1), and ideal-self ratings did not predict observer impressions above and beyond actual personality. (PsycINFO Database Record (c) 2016 APA, all rights reserved).

3. The Big Five personality dimensions and job performance: A meta-analysis Emotional Stability, Agreeableness, Conscientiousness, and Openness to Experience) to three job performance criteria (job proficiency, training proficiency, and personnel data) for five occupational groups (professionals, police, managers, sales, and skilled/semi-skilled). Results indicated that one dimension of personality, Conscientiousness, showed consistent relations with all job performance criteria for all occupational groups. For the remaining personality dimensions, the estimated true score correlations varied by occupational group and criterion type. Extraversion was a valid predictor for two occupations involving social interaction, managers, and sales (across criterion types). Also, both Openness to Experience and Extraversion were valid predictors of the training proficiency criterion (across occupations). Other personality dimensions were also found to be valid predictors for some occupations and some criterion types, but the magnitude of the estimated true score correlations was small ($\rho < .10$). Overall, the results illustrate the benefits of using the 5-factor 5 model of personality to accumulate and communicate empirical findings. The findings have numerous implications for research and practice in PERSONNEL PSYCHOLOGY, especially in the subfields of personnel selection, training and development, and performance appraisal.

4. Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance: Investigated the moderating role of autonomy on the relationships between the Big Five personality dimensions and supervisor ratings of job performance. Based on data from 146 managers, results indicate that 2 dimensions of personality, Conscientiousness (r.25) and Extraversion (= r.14), were significantly related to job performance. Consistent with expectations, the validity of Conscientiousness and Extraversion was greater for managers in jobs high in autonomy compared with those in jobs low in autonomy. The validity of Agreeableness was also higher in high-autonomy jobs compared with low-autonomy ones, but the correlation was negative. These findings suggest that degree of autonomy in the job moderates the validity of at least some personality predictors. Implications for future research are noted. (PsycINFO Database Record (c) 2016 APA, all rights reserved).

5. The right relationship is everything: Linking personality preferences to managerial behaviours: Individual differences and personality factors have reemerged as some of the more important research topics in the applied organizational sciences. With the increasing prevalence of executive coaching and the use of personality assessments, more research needs to be done on the impact of personality variables on managerial behaviours in the workplace. The following study provides an applied analysis of personality preferences and behavioural ratings collected for a developmental multirotor feedback intervention based on 343 senior managers and others in a research-driven global health services organization. Results revealed modest personality behaviour relationships, many of which were consistent with Myers-Briggs Type Indicator theory and research; differences by observer perspective were also evident. Implications for HRD practices are discussed. © 2000 by Jossey-Bass, A Publishing Unit of John Wiley & Sons, Inc.

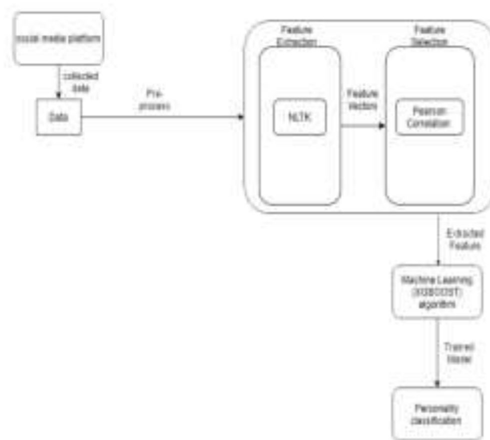
3.METHODOLOGY

Existing System:

The ability to predict personality has implications in many areas. Existing research has shown connections between personality traits and success in both professional and personal relationships. Social media tools that seek to support these relationships could benefit from personality insights. Additionally, previous work on personality and interfaces showed that users are more receptive to and have greater trust in interfaces and information that is presented from the perspective of their own personality features (i.e., introverts prefer messages presented from an introvert's perspective). If a user's personality can be predicted from their social media profile, online marketing and applications can use this to personalize their message and its presentation.

Proposed System:

In this paper author is asking to predict human personality by taking 5 features such as Openness, Agreeable, Neuroticism, Extroversion, and Conscientious and this feature can be identified by social media dataset called 'Twitter Profile'. Openness means intelligent peoples who express their view in bold or open manner. This user expression can be identified by analysing his twitter profile and twitter messages, if person is intelligent then he will use open (open words also called as swear words) or bold words in his tweets. By looking for such words we can identify this person as Openness personality. LIWC dictionary contains all open or swear words by applying this dictionary on tweets messages we can predict Openness personality score. If predicted score > 0.1 then this person will put under this category. Agreeable means peoples who use words such as 'am, will have and this words also refers as ARTICLES or AUXILIARY VERBS' etc will come in this category. MRC dictionary contains all words of this categories and by applying this dictionary on user's tweets we can predict person category as agreeable. Neuroticism means peoples in this category is consider as sentiment or emotion, peoples who use words such as 'ugly, nasty, sad' etc will come under this category. By looking for such words in tweets we can predict score of this category. Extroversion means peoples of this category are friendly and person who has many numbers of friends or followers or following in twitter profile will comes under this category. Conscientious means peoples who express hard working ideas in their post will come under this category. So, by analysing above 5 features O (openness), C (Conscientious), E (Extroversion), A (Agreeable), N (Neuroticism) from twitter profile and post we can predict personality of a person. 2 We will find average of each feature from tweets and then apply Pearson Correlation formula to get scores for all five features. If score > 0.1 for any feature, then person belongs to that category. If person has 0.1 value for more than 1 features, then that person personality belongs to that many categories. For example, same person can be predicted as openness and conscientious etc. All features' values we will apply using SVM, Random Forest, Naïve Bayes & Logistic Regression algorithms to calculate accuracy of dataset and algorithms.

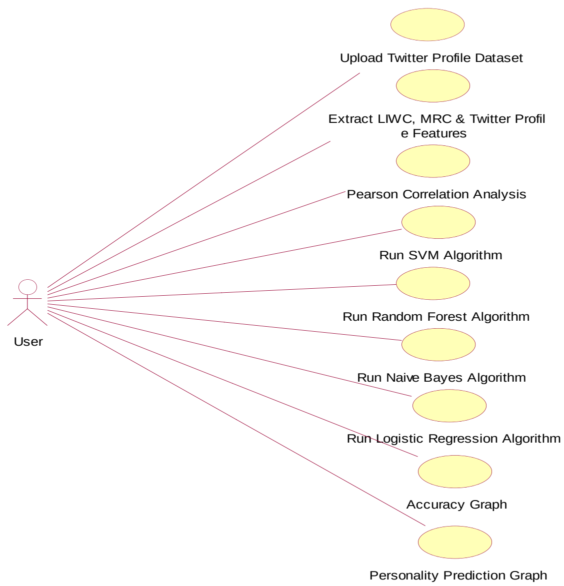


UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modelling language in the field of object-oriented software engineering. The standard is managed and was created by the Object Management Group. The goal is for UML to become a common language for creating models of object-oriented computer software. In its current form, UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML. The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artefacts of software systems, as well as for business modelling and other non-software systems. The UML represents a collection of best engineering practices that have proven successful in the modelling of large and complex systems. The UML is a very important part of developing object- oriented software and the software development process. UML uses mostly graphical notations to express the design of software projects.

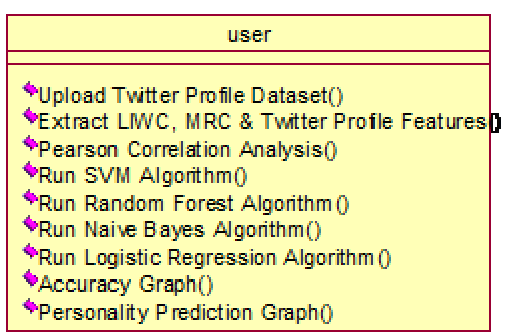
USE CASE DIAGRAM

A use case diagram in the Unified Modelling Language (UML) is a type of behavioural diagram defined by and created from a Use-case analysis.



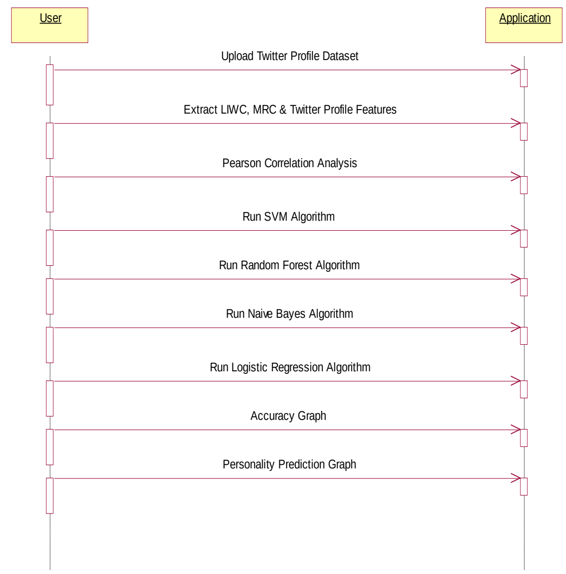
CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modelling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



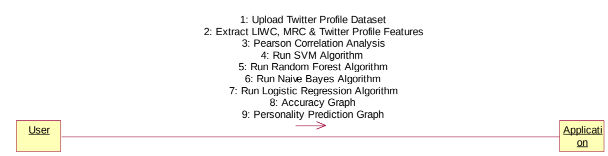
SEQUENCE DIAGRAM:

A sequence diagram in Unified Modelling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.



COLLABORATION DIAGRAM:

Collaboration diagrams are particularly useful during the design and analysis phase of software development to visualize and understand the communication flow between components or objects in a system.



SYSTEM STUDY

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and the business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are.

- ♦ ECONOMICAL FEASIBILITY
- ♦ TECHNICAL FEASIBILITY
- ♦ SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impactthat the system will have on the organization. Theamount of funds that the company can pour into the research and development of the system is limited. The expenditure must be justified. Thus, the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products hadto be purchased.

TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. No system developed must not have a high demand on the available technical resources. This will lead to high demands for the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is thefinal user of the system.

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail unacceptably. There are various types of tests.

TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional testing

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

System Testing

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a test in which the software tester has knowledge of the inner workings, structure, and language of the software, or at least its purpose. It is the purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, like most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a test in which the software under test is treated, as a black box. you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works.

Unit Testing

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

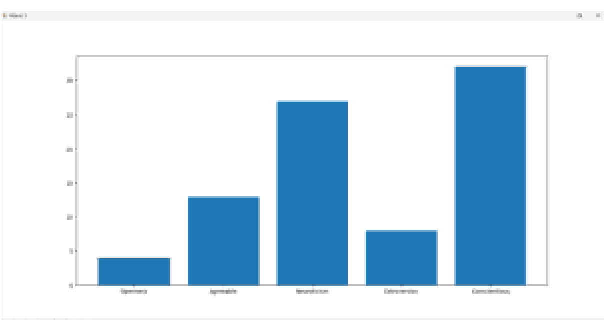
Integration testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects. The task of the integration test is to check that components or software applications, e.g., components in a software system or – one step up – software applications at the company level – interact without error.

Acceptance Testing

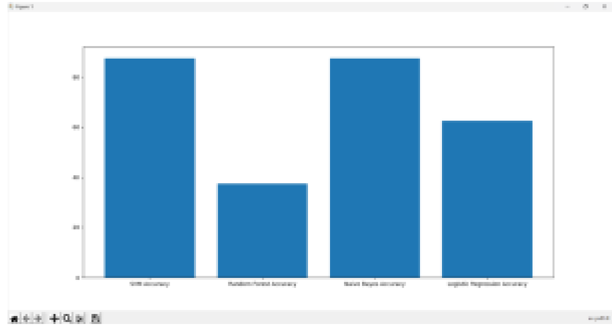
User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

4. EXPERIMENTAL OUTCOME



• The above figure shows the big 5 personality model.
• Where here Conscientiousness has more accuracy than other 4 personality model.
• By this we can say that more people personality come under Conscientiousness

This concise output design provides a structured overview of a paper on predicting personality from Twitter data.



In above graph x-axis represents algorithm name and y-axis represents accuracy of that algorithm. Now click on 'Personality Prediction Graph' to get number of peoples in each category graph.

5.CONCLUSION AND FUTURE SCOPE:

Conclusion:

In this paper, we have shown that a users' Big Five personality traits can be predicted from the public information they share on Twitter. Our subjects completed a personality test and through the Twitter API, we collected publicly accessible information from their profiles.

After processing this data, we found many small correlations in the data. Using the profile data as a feature set, we were able to train two machine learning algorithms - ZeroR and Gaussian Processes - to predict scores on each of the five personality traits to within 11% - 18% of them

actual value. With the ability to guess a user's personality traits, many opportunities are opened for personalizing interfaces and information. We discussed some of these opportunities for marketing and interface design above. However, there is much work to be pursued in this area.

One area that deserves attention is the connection between personality and the actual social network. We considered two structural features - number of friends and network density but we did not look at personality scores between friends. Understanding the connections between personality, tie strength, trust, and other related factors is an open space for research. By improving our knowledge of these relationships, we can begin to answer more sophisticated questions about how to present trusted, socially relevant, and well-presented information to users.

FUTURE SCOPE:

Predicting personality from Twitter data holds promise for future enhancements in various domains. Potential improvements could include refining data collection techniques, enhancing machine learning models, and applying personality insights to personalize user experiences, including social media interactions, content recommendations, and interface design.

6. REFERENCES:

[1] General inquirer, 1997. http://www.wjh.harvard.edu/ inquirer/.

[2] A. Acar and M. Polonsky. Online Social Networks and Insights into Marketing Communications. Journal of Internet Commerce, 6(4):55– 72, 2008.

[3] M. Back, J. Stopfer, S. Vazire, S. Gaddis, S. Schmukle, B. Egloff, and S. Gosling. Facebook Profiles Reflect Actual Personality, Not SelfIdealization. Psychological Science, 21(3):372, 2010.

[4] M. Barrick and M. Mount. The Big Five personality dimensions and job performance: A meta-analysis. Personnel psychology, 44(1):1–26, 1991.

[5] M. Barrick and M. Mount. Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. Journal of Applied Psychology, 78(1):111–118, 1993.

[6] S. Berr, A. Church, and J. Waclawski. The right relationship is everything: Linking personality preferences to managerial behaviors. Human Resource Development Quarterly, 11(2):133–157, 2000.

[7] W.-P. Brinkman and N. Fine. Towards customized emotional design: an explorative study of user personality and user interface skin preferences. In EACE '05: Proceedings of the 2005 annual conference on European association of cognitive ergonomics, pages 107–114.
University of Athens, 2005.

[8] T. Chamorro-Premuzic. Personality and Romantic Relationships, volume Personality and Individual Differences. Blackwell Publishing, 2007.

[9] B. De Raad. The Big Five personality factors: The psycholexical approach to personality. Hogrefe & Huber G"ottingen, 2000.

[10] J. Digman. Personality structure: Emergence of the five-factor model. Annual review of psychology, 41(1):417–440, 1990.

[11] S. Dollinger. Research Note: Personality and Music Preference: Extraversion and
Excitement Seeking or Openness to Experience? Psychology of Music, 21(1):73, 1993.

[12] T. DuBois, J. Golbeck, J. Kleint, and A. Srinivasan. Improving Recommendation Accuracy
by Clustering Social Networks with Trust. In Recommender Systems & the Social Web, 2009.

[13] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In Proceedings of
the 27th international conference on Human factors in computing systems, pages 211–220.
ACM New York, NY, USA, 2009.

[14] J. Golbeck. Computing and Applying Trust in Web-based Social Networks. PhD thesis,
University of Maryland, College Park, MD, USA, April 2005.

[15] J. Golbeck, C. Robles, and K. Turner. Predicting personality with social media. In Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems, CHI EA '11, pages 253–262, New York, NY, USA, 2011. ACM.34

[16] L. Goldberg. From Ace to Zombie: Some explorations in the language of personality.
Advances in personality assessment, 1:203–234, 1982.

[17] S. D. Gosling, C. J. Sandy, and J. Potter. Personalities of self-identified dog people and cat people. Anthrozoos: A Multidisciplinary Journal of The Interactions of People & Animals, 23:213–222(10), September 2010.

[18] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten. The WEKA data mining software: An update. ACM SIGKDD Explorations Newsletter, 11(1):10–18, 2009.

[19] C. Hansen and R. Hansen. Constructing personality and social reality through music: Individual differences among fans of punk and heavy metal music. Journal of broadcasting & electronic media, 35(3):335– 350, 1991.

[20] G. Hodgkinson and J. Ford. International Review of Industrial and Organizational Psychology, 2008. Wiley-Interscience, 2008.