# Beyond Facebook Personality Prediction:

## A Multidisciplinary Approach
## to Predicting Social Media Users' Personality

Carrie Solinger[1], Leanne Hirshfield[1], Stuart Hirshfield[2],
Rachel Friendman[2], and Christopher Lepre[2]

[1] Department of Mass Communications, S.I. Newhouse School of Public Communications,
Syracuse University, Syracuse, NY 13210
`{casoling,lmhirshf}@syr.edu`
[2] Department of Computer Science, Hamilton College, Clinton, NY
`{shirshf,rlfriendm,cleper}@hamilton.edu`

**Abstract.** We investigate creating a predictive model that increases accuracy in personality prediction of social media and social network site users through a multidisciplinary pilot analysis. We present a novel method for increasing personality prediction accuracy of Facebook users. We discuss an experiment that combines natural language processing and machine learning methods, as well as the Big Five Personality and other cognitive psychology metrics and scales. Our machine learning predictive model showed promising results in personality prediction accuracy of three personality traits. However, the results indicate that more research and further data collection will improve prediction accuracies.

**Keywords:** Big Five, personality traits, propensity to trust, trust propensity, need for cognition, Facebook, social media, social networking sites, SNS.

## 1 Introduction

Accurately predicting social media users' behaviors and activity on social networking sites (SNS), such as Facebook and Twitter, is a quickly expanding research domain; this is no surprise given the exponential growth of social media and SNS users. Pew Research Center reported that 72% of online adults are active users of social networking sites [1]. This figure is more than double the 35% reported in 2009 [2]. Additionally, of the 72% adults with online profiles, 67% of these individuals use Facebook [1].

Existing social media and SNS personality prediction research is primarily within the context of a single discipline, such as human-computer interaction, social computing, and behavior modeling [3-6]. While each of these prior studies provides valuable insights, we believe there is greater benefit in applying a multifaceted approach to increase prediction accuracy. Specifically, inclusion of additional components from

computer science, computational linguistics, cognitive psychology, and information science would be advantageous.

The primary goals of our study were to increase accuracy in social media and SNS user personality prediction, and determine his or her susceptibility to (social) influence by others in social media and SNS environments. As such, our research study implemented a multidisciplinary method for higher personality predication accuracy, and broader-based applicability in understanding how users of social media and SNS act and behave based on the individual user's representation of his or her own social identity and social presence within the social media and SNS domains. In particular, our study differed from prior studies via application of sentiment analysis and machine learning (ML) to conduct contextual analysis though classifiers. Additionally, cognitive psychology metrics and scales were utilized through survey deployment.

## 2    Background and Relevant Literature

### 2.1    Personality

As social media and SNS user populations have grown, research studies within these areas have similarly expanded. In particular, the high usage level of the Facebook social network has garnered attention in the personality prediction research domain. Many of these research studies applied the Big Five factors (of personality traits) model to Facebook profile information and usage in assessing personality prediction accuracy. The Big Five, also known as the Five-Factor Model, consists of five broad personality dimensions used to describe the human personality [7]. Often collectively referred to as OCEAN, these factors are: openness, conscientiousness, extraversion, agreeableness, and neuroticism. Table 1 illustrates the five personality traits and values describing HIGH and LOW scorers for each.

Research conducted using the Big Five as the basis for personality prediction, as applied to Facebook, indicated correlations between some of the personality factors and publically available Facebook profile information [6, 8]. Significant findings in these studies included prediction accuracy of each of the five factors within 11% of the user's survey based values [8]. Additionally, these studies reported the highest prediction accuracy for *extraversion*, and the lowest prediction accuracy for *agreeableness* [6, 8].

**Table 1.** The Big Five (Factor) Model descriptions

| Personality Trait | HIGH scorers tend to be: | LOW scorers tend to be: |
|---|---|---|
| *Openness* | Creative, imaginative, unconventional, intellectually curious, open to new experiences. | Conventional, traditional, conservative, prefer familiarity, skeptical. |
| *Conscientiousness* | Organized, self-disciplined, efficient, hardworking, reliable, neat, systematic, thorough. | Easy-going, spontaneous, careless. |
| *Extraversion* | Energetic, enthusiastic, talkative, assertive, gregarious. | Quiet, low-key, independent, prefer time alone. |
| *Agreeableness* | Kind, warm, considerate, cooperative, trustworthy, trusting, friendly, sympathetic. | Competitive, analytical, skeptical, suspicious, unfriendly, detached. |
| *Neuroticism* | Sensitive, anxious, moody, self-conscious, perfectionist, more reactive to stress. | Calm, even-tempered, less likely to be affected by stressors. |

## 2.2    Propensity to Trust

Trust is a complex phenomenon and considered by many different research disciplines to be primarily situational within the social context [9]. In relation to personality, trust is often included as a facet of the Big Five agreeableness dimension. However, some scholars dispute the notion of trust as a momentary state. Instead, trust is reinterpreted as a *propensity to trust,* and therefore viewed as an underlying aspect of personality [9, 10]. The Propensity to Trust Scale was created to more accurately measure trust related behavior. This assessment is an 8-item Likert-like scale that estimates an individual's wiliness to trust other people [11].

## 2.3    Need for Cognition

Need for cognition is the overall tendency for individuals to engage in and enjoy intensive thinking [12]. An 18-item Likert-like scale was created to measure an individual's need for cognition [13]. The assessment indicates characteristics about individuals with HIGH or LOW need for cognition. Individuals with high need for cognition are more likely to receive well a message that emphasizes cognitive information rather than emotion. Conversely, individuals low in need for cognition tend to form opinions based on more simple cues such as the credibility or attractiveness of the message source. Additionally, higher scoring need for cognition individuals tend to have stronger opinions than their lower scoring counterparts [14].

**Table 2.** 8-item Propensity to Trust Scale statements

| Item Statements |
|---|
| Most people can be counted on to do what they say they will do. |
| I tend to trust people, even those whom I have just met for the first time. |
| Unless you remain alert, someone will soon take advantage of you. |
| Most people would tell a lie if they could gain by it. |
| My typical approach is to be cautious with people until they have demonstrated their trustworthiness. |
| I usually give acquaintances the benefit of the doubt if they do something that seems selfish. |
| Most people pretend to be more honest than they really are. |
| I believe that most people are generally trustworthy. |

## 3    Experiment

### 3.1    Participants

Our pilot analysis sample population consists of 20 college students. The majority of the participants (90%) are in the 18-24 age group. The remaining two participants are in the 25-34 and 35-54 age groups. Out the 20 participants, 15 are female and 5 are male. Participants were required to have existing Facebook accounts.

### 3.2    Facebook Application Access

Each participant was required to download a program while logged into his or her own Facebook account. The program was built into the existing Facebook application interface (API), and collected the participants' account data. Data collected via this application included profile bios, status updates, photos, and number of "Friends" (account-to-account connections). Data collection was limited to a two-year timespan, from June 2011 – June 2013.

### 3.3    Surveys

Participants used an online survey tool to complete three self-report personality-based surveys: Big Five personality inventory, propensity to trust assessment, and need for cognition assessment. The three personality-based surveys measured seven dimensions. The Big Five personality inventory measured *openness*, *conscientiousness*, *extraversion*, *agreeableness*, and *neuroticism*. The *propensity to trust* and *need for cognition* dimensions were measured by its respective assessment. Participants' responses were scored and assigned values of either HIGH or LOW.

**Table 3.** Need for Cognition Scale Short Form

| Item Wording |
|---|
| I would prefer complex to simple problems. |
| I like to have the responsibility of handling a situation that requires a lot of thinking. |
| Thinking is not my idea of fun. |
| I would rather do something that requires little thought than something that is sure to challenge my thinking abilities. |
| I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something. |
| I find satisfaction in deliberating hard and for long hours. |
| I only think as hard as I have to. |
| I prefer to think about small, daily projects to long-term ones. |
| I like tasks that require little thought once I've learned them. |
| The idea of relying on thought to make my way to the top appeals to me. |
| I really enjoy a task that involves coming up with new solutions to problems. |
| Learning new ways to think doesn't excite me very much. |
| I prefer my life to be filled with puzzles that I must solve. |
| The notion of thinking abstractly is appealing to me. |
| I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought. |
| I feel relief rather than satisfaction after completing a task that required a lot of mental effort. |
| It's enough for me that something gets the job done; I don't care how or why it works. |
| I usually end up deliberating about issues even when they do not affect me personally. |

### 3.4    Data Preprocessing and Machine Learning Analysis

Data analysis consisted of three-primary phases. In the first phase, participant responses from each of the three surveys were converted into seven different predictor variables, which represented personality, propensity to trust, and need for cognition, respectively. The participants' predictor variables were ranked as either high or low.

For the second phase, an open-source natural language processing tool, Natural Language Toolkit, was used to create and apply a sentiment analysis classifier. This classifier was applied to the contextual data collected from the participants' Facebook account, which included user-generated status updates and user-generated comments. Statistical data output was converted to percentages for machine learning (ML) preparation. Lastly, an open-source data mining with ML application, WEKA [15], was used to build a neural network classifier. The classifier consisted of 10 hidden layers with a learning rate of .1. A three-fold cross validation was used to test the accuracy of our models.

## 4    Results and Analysis

### 4.1    Machine Learning Classifier

The most significant results from the initial application of our ML classifier indicates a personality prediction of 75% accuracy in *need for cognition*, and 65% accuracy in *agreeableness* of Facebook users. The latter result is significant as prior studies indicated a low accuracy prediction of this personality trait [6, 8]. Additionally, we were able to replicate a similar level of *extraversion* prediction accuracy with that of prior studies [6, 8]. While our pilot study consisted of only 20 Facebook users, we anticipate these prediction accuracies to improve with further data collection.

**Table 4.** Personality prediction accuracy results

| Personality Dimensions | Prediction Accuracy |
|---|---|
| Openness | 50% |
| Conscientiousness | 40% |
| *Extraversion* | *65%* |
| *Agreeableness* | *65%* |
| Neuroticism | 55% |
| Propensity to Trust | 55% |
| *Need for Cognition* | *75%* |

**Table 5.** Survey results for the four participants that passed along the Kony video

| Sent Kony | Openness | Conscientiousness | Extraversion | Agreeableness | Neuroticism | Propensity to Trust | Need for cognition |
|-----------|----------|-------------------|--------------|---------------|-------------|---------------------|--------------------|
| YES | HIGH | HIGH | HIGH | HIGH | LOW | HIGH | HIGH |
| YES | LOW | LOW | LOW | LOW | HIGH | HIGH | LOW |
| YES | LOW | LOW | LOW | HIGH | LOW | HIGH | LOW |
| YES | LOW | LOW | LOW | LOW | HIGH | HIGH | HIGH |

## 4.2    Kony 2012

Of the 20 subjects, only four subjects passed along a viral event (the Kony video). These numbers are too small to produce significant findings, but we did look at the personality, propensity to trust, and need for cognition survey results for each of these four participants to see if there were any notable trends. Table 5 shows the survey results for these four participants. It is interesting to note that all of these subjects had a high propensity to trust. And three of the four participants had low levels of openness and conscientiousness.

Although we can not draw any concrete conclusions from such a small sample, the trends seen in the data from these four subjects suggests that further research should explore the ties between propensity to trust, personality type, and whether or not that individual will be likely to send a viral event.

## 5    Conclusion

In this paper, we discussed and presented relevant Facebook personality prediction studies, and the benefit of including additional cognitive psychology metrics in a multidisciplinary approach to increase Facebook personality prediction accuracies. We demonstrated that the inclusion of additional personality dimensions, specifically the need for cognition and agreeableness, may be predicted with a high degree of accuracy. Although our pilot analysis shows promising results, future work to explore additional factors and a larger participant pool is needed to achieve a more stable machine learning predictive personality model.

# References

1. Brenner, J., Smith, A.: 72% of online adults are social networking site users (2013)
2. Lenhart, A.: Adults and social network websites (2009)
3. Amichai-Hamburger, Y., Vinitzky, G.: Social network use and personality. Computers in Human Behavior 26(6), 1289–1295 (2010)
4. Quercia, D., et al.: Our twitter profiles, our selves: Predicting personality with twitter. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (passat) and 2011 IEEE Third International Conference on Social Computing (socialcom). IEEE (2011)
5. Golbeck, J., et al.: Predicting personality from twitter. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (passat) and 2011 IEEE Third International Conference on Social Computing (socialcom). IEEE (2011)
6. Bachrach, Y., et al.: Personality and patterns of Facebook usage. In: Proceedings of the 3rd Annual ACM Web Science Conference. ACM (2012)
7. Costa, P.T., McCrae, R.R.: The revised neo personality inventory (neo-pi-r). In: The SAGE Handbook of Personality Theory and Assessment, vol. 2, pp. 179–198 (2008)
8. Golbeck, J., Robles, C., Turner, R.: Predicting personality with social media. In: CHI 2011 Extended Abstracts on Human Factors in Computing Systems 2011, pp. 253–262. ACM, Vancouver (2011)
9. Evans, A.M., Revelle, W.: Survey and behavioral measurements of interpersonal trust. Journal of Research in Personality 42(6), 1585–1593 (2008)
10. Grabner-Kräuter, S., Kaluscha, E.A.: Empirical research in on-line trust: a review and critical assessment. Journal of Human-Computer Studies 58(6), 783–812 (2003)
11. McShane, S., von Glinow, M.: Propensity to trust scale. In: Organizational Behavior: Essentials 2008. McGraw-Hill (2008)
12. Cacioppo, J.T., Petty, R.E.: The need for cognition. Journal of Personality and Social Psychology 42(1), 116 (1982)
13. Cacioppo, J.T., et al.: Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. Psychological Bulletin 119(2), 197 (1996)
14. Petty, R.E., et al.: The need for cognition. In: Handbook of Individual Differences in Social Behavior, pp. 318–329 (2009)
15. Hall, M., et al.: The WEKA Data Mining Software: An Update. SIGKDD Explorations 11(1) (2009)