Stanford University CS 109 Autumn Project

Big Five Twitter Personality Traits Project

Alejandrina G.R., alegre@cs.stanford.edu

https://big5project.alejandrina.me

December 2, 2019

Abstract

Internet and social media are a big part of people's lives nowadays. Previous research indicates that problematic Internet usage is associated with the personality traits of introversion and neuroticism [1]. Further, researchers say there is a strong connection between a user's behaviour online and their personality. [2].

While online behavior can take different forms, I decided to pursue the question of whether personality traits from the Big-Five Factor Model can be estimated *just* from social media text based status updates of a user account, specifically Twitter users, with text classification. Since personality is a multidimensional analysis and social media activity is often shallow and one-dimensional, I hypothesized that social media statuses analyzed with text classification algorithms can only give a glimpse of a user's online personality traits and not a fair analysis of their actual personality traits off the internet.

Keywords

Twitter, personality, text classification, big five factor

1 Introduction

The ability of CS techniques to be able to solve social science problems and questions lead me to inquire about how online social media expression can predict a user's personality.

On the social science side of this project, in order to analyze personality, I decided to use the Big-Five Factor Model recognized by the American Psychological Association and which is the base of thousands of peer-reviewed papers on personality. While personality is a multidimensional analysis, the Big-Five Factor Model focuses and assesses the following five personality traits: Extraversion, proneness to socialize and to have positive emotions; Agreeableness, inclination of being sympathetic and cooperative; Conscientiousness, tendency to be organized and hard-working; Neuroticism, sensitivity and proclivity of experiencing psychological distress; and Openness, disposition of considering alternative views [3].

On the statistical and computer science side of the project, I wrote Näive Bayes text classification algorithms in order to predict the probability of each personality trait being present in an specific social media user.

Not only did I just write the algorithms to analyze text and estimate the personality traits, I also built a complete system around it in order to make my proof of concept project available on my website for anyone curious to try it out: https://big5project.alejandrina.me

2 Materials & Methods

For the text classification models, I wrote Näive Bayes classification algorithms from scratch using the Numpy[1] and Pandas[2] python libraries.

[1] https://numpy.org
[2] https://pandas.pydata.org

For the backend, I used an AWS[3] instance server with an mySQL[4] database and end-to-end encryption with Let's Encrypt[5]. For the frontend, I used Flask[6] and the Twitter API[7] to build a multithreaded web application that would take any public Twitter profile handle @username.

Corpus used to train the text classification models was originally created by a team of researchers from the University of Cambridge Psychometrics Centre [7,8] with a Facebook application. The corpus includes anonymous Facebook statuses in raw text with associated gold standard labels (classes and scores). Since in the corpus each proper name has been replaced with the string *PROPNAME*, then Tweets retrieved were cleaned as well to replace user handle mentions with *PROPNAME*. Corpus gold standard labels included:

- sEXT extraversion (score)
- sNEU neuroticism (score)
- sAGR agreableness (score)
- sCON conscientiousness (score)
- sOPN openness (score)

- cEXT extraversion (y=extravert,n=shy)
- cNEU neuroticism (y=neurotic,n=secure)
- cAGR agreableness (y=friendly,n=uncooperative)
- cCON conscientiousness (y=precise,n=careless)
- cOPN openness (y=insightful,n=unimaginative)

The corpus was divided into two sets, 90% for training set and 10% for testing set, by randomly picking the corresponding percent of rows from the dataset with no duplication. Trained model achieved an accuracy of predicting personality by 71% with the testing set it had not encountered before. The personality trait with the maximum accuracy was Openness.

3 Results

My model trained model a mas accuracy of 75%. I was able to collect data of users who tried my project and who had previously taken an official Big Five Factor personality test. Users openly shared with me their results of both for me to compare results from my project model's accuracy and the official Big Five Factor personality test.

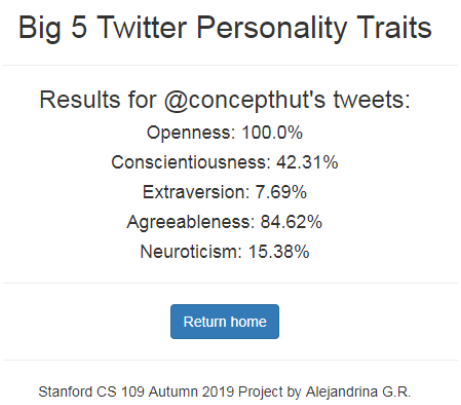

Figure 1: Image shared by Twitter user @ConceptHut

Twitter user @ConceptHut shared with me the following results from using my classification model (Fig. 1) and their results of a previously taken Big Five Factor personality test[9,10]:

> (Values in percentiles)
> Openness: 94st
> Conscientiousness: 28st (bad few years...)
> Extroversion: 96st
> Agreeableness: 61st
> Neuroticism: 20th

Further, Twitter user @LennyBruce14 shared the following results from my project vs the results from the Big Five Factor test from Open Psychometrics[11,12]:

> Extroversion: 30% / 41%
> Conscientiousness: 34% / 31%
> Agreeableness: 54% / 2%
> Openness: 89% / 70%
> Neuroticism: 34% / 9% (Open Psychometrics result is given inverted, and was 91

A third Twitter user, @EkmoreGG, shared an image comparing the my project's results (Fig. 2) vs the Big Factor Five test results (Fig. 3) form Understanding Myself[13,14]. Last but not

[3] https://aws.amazon.com
[4] https://www.mysql.com
[5] https://letsencrypt.org
[6] https://www.palletsprojects.com/p/flask/
[7] https://developer.twitter.com
[7] http://myPersonality.org
[8] https://www.psychometrics.cam.ac.uk

[9] https://twitter.com/ConceptHut/status/1201059118674595840?s=20
[10] https://www.understandmyself.com
[11] http://openpsychometrics.org
[12] https://twitter.com/LennyBruce14/status/1201140491942793216?s=20
[13] https://www.understandmyself.com
[14] https://twitter.com/EkmoreGG/status/1201345067132153857?s=20

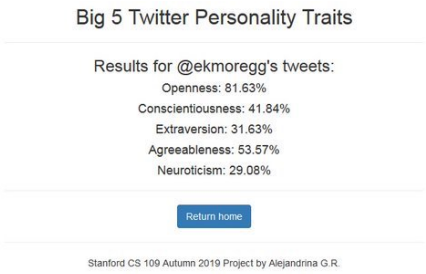

Figure 2: Shared by Twitter user @EkmoreGG

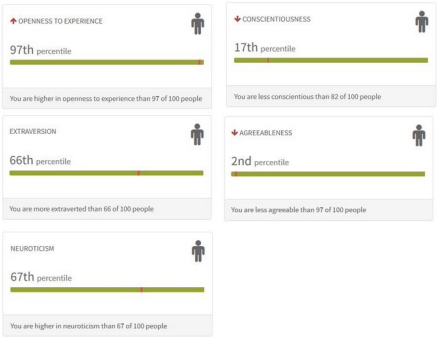

Figure 3: Shared by Twitter user @EkmoreGG

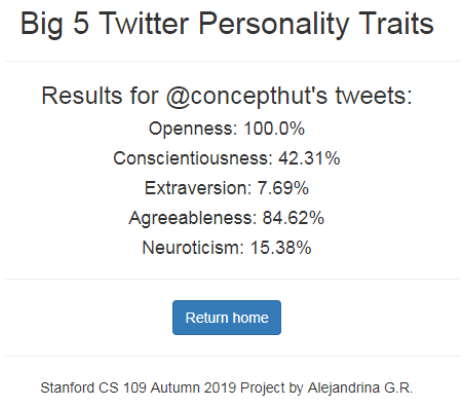

Figure 4: Results of @Jaqaliah from my model

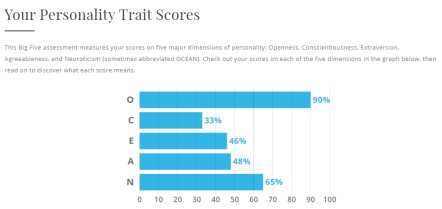

Figure 5: Results of @Jaqaliah from an online Big Five Factor test

least, Twitter user @Jaqaliah shared their results (Fig. 4, 5) for comparison[15,16].

4 Discussion

While I was able to collect some data from some users for simple comparison, I was not able to collect enough to do a thorough analysis of the results and find the statistical significance. However, it is surprising that my model's results for the trait of Openness was the most closely accurate to the results from the standard Big Five Factor tests.

Conclusions/Future Work

While this was a proof of concept project, it could be improved further. The text classification algorithm can be improved by taking into account how the five personality traits relate to each other along with the text of the user statuses in order to improve the model's accuracy above 71%. A higher number of participants would be needed. My model would also need to be tested on individuals that have taken the same (and not different) online Big Five Factor test to analyze standard results.

Source code: https://github.com/agonzalezreyes/109Project

Acknowledgements

Thanks to the CS 109 Teaching Staff. To all the Twitter users that tried my program when I launched it on Nov. 30th, 2019. Thanks to Stripe CEO, Patrick Collins, and Siri co-founder, Adam Cheyer, for recognizing the insightfulness of this proof of concept project right after the launch.

References

[1] Susan Ebeling-Witte, Michael L Frank, and David Lester. Shyness, internet use, and personality. *CyberPsychology & Behavior*, 10(5):713–716, 2007.

[2] Yair Amichai-Hamburger and Gideon Vinitzky. Social network use and personality. *Computers in human behavior*, 26(6):1289–1295, 2010.

[3] Lewis R Goldberg. An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216, 1990.

[15] https://twitter.com/Jaqaliah/status/1201095800119410688?s=20
[16] https://twitter.com/Jaqaliah/status/1201585845817532416?s=20