

Predicting personality based on social media posts

Lisa Gotzian

Received: March 15, 2019

Abstract Personality prediction from publicly available data such as social media behavior and social media posts has gained immense interest in research. Knowing a person's personality based on the Big 5 has vast implications for one's privacy. Therefore, information on the Big 5 is heavily restricted. Here, a dataset with MBTI personalities is analyzed to find resemblance with the Five Factor Model and to determine if the more accessible MBTI personality types pose a similar threat on privacy. The data comes from a personality forum. An SVM with a number of psycholinguistic features such as a word-emotion lexicon reaches 54% accuracy, beating current results in research. However, the dataset has to be treated carefully as it was collected from personality enthusiasts and might facilitate personality prediction.

Keywords MBTI · Personality Prediction · Text Analysis · Support Vector Machines

1 Introduction

One of the most important concepts in psychology is an individual's personality. As it is stable over the course of 20 years [16], its effect on all kinds of different behaviors and concepts such as job performance [27], happiness [20] and marital satisfaction [9] has been evaluated. The insights of personality research are valuable for any research conducted around humans such as organizational psychology, human-machine interaction or customer management just to name a few.

The most famous model to describe personality is the **Five Factor Model**, the Big 5 [16]. The model states how each personality can be described using 5

Lisa Gotzian
M. Sc. *Management & Data Science*
E-mail: lisa.gotzian@stud.leuphana.de

bipolar dimensions: *openness to new experience*, *conscientiousness*, *extraversion*, *agreeableness* and *neuroticism*.

The model has been cross-validated and is broadly accepted. It is usually measured using tests like the NEO PI-R [16]. These 60 min long tests are time-consuming and require lots of efforts when recruiting participants. In addition, especially personality data is strongly protected as this shares very private information [11]. Thus, datasets with information on the Big 5 are hard to accumulate and are potentially dangerous to work with.

In contrast to the Big 5, the **Myers-Briggs Type Indicator** evaluates personality on 4 dimensions: *Introversion/Extraversion* (how one gains energy), *Sensing/iNtuition* (how one processes information), *Thinking/Feeling* (how one makes decisions), and *Judging/Perceiving* (how one presents herself or himself to the outside world) [18]. The combination of all 4 dimensions yields 16 different personality types based on the works by Carl Jung [8]. Though these dimensions correlate with 4 of the 5 dimensions of the Big 5 [15], it lacks the last dimension, *neuroticism*. The MBTI however is not as broadly accepted in the scientific community as it fails to provide a valid model with clearly distinguishable dimensions [2, 15].

For research, the MBTI provides use nevertheless: first, especially in the corporate world, the MBTI enjoys great popularity. It is much easier to retrieve correctly labeled data, e.g. from Twitter [22] or from reddit [4]. In both cases, users deliberately show their personality type and contribute to datasets with several thousand entries.

Second, due to its closeness to the Big 5, the MBTI might be able to predict certain Big 5 features, in return leading to usable results but with the advantage of more accessible data. Datasets on personality especially involving the Big 5 are hard to get ahold of due to privacy issues or maintenance reasons [26].

Recently, machine learning researchers have started to infer the Big 5 from more indirect data. Instead of elaborate tests, they predicted personality based on text [14, 21] like Twitter posts [6, 23], other social media posts [24] and emails [19]. Personality can also be inferred from general social media behavior as likes, interactions and other digital records [12, 10, 28]. The results come with an accuracy of 0.4-0.5 and up to 0.8 for individual dimensions, beating even the ratings by close family members [28].

This paper intends to (1) predict MBTI types based on social network posts as done with the Big 5 to then (2) determine if the MBTI is a valid measure for text data. Based on this, one could derive the user's Big 5 which would pose a similar threat on privacy as the Big 5 itself.

2 The Data

The data consists of 8675 rows of text, each with 50 posts or comments including the MBTI type of said person. It has been scraped from the "Personality

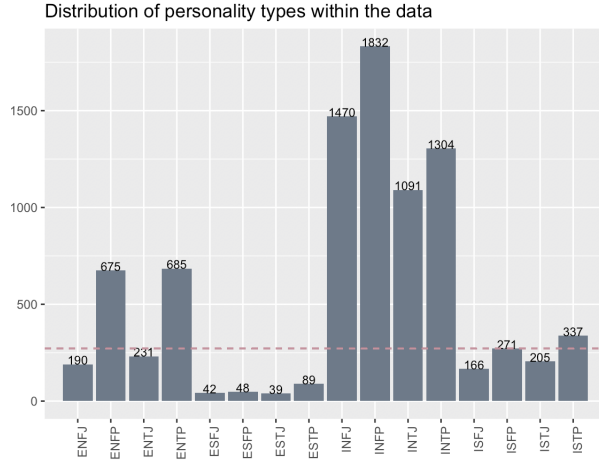


Fig. 1: Histogram of all personality types including a threshold to indicate types that occurred less than half of the expected times.

Cafe forum”¹ and made publicly available². The forum is used and maintained by personality enthusiasts which is why the users provide their MBTI type as part of their profile and thus allow for correctly labeled data.

The different personality types are distributed unevenly in the data, see figure 1. Therefore, this analysis includes a model without types that occurred less than half of the expected rate for 16 personality types $\frac{1}{16 \times 2}$, 3.1% or 271 observations. This more robust model uses 7 out of 16 personality types for a more stable estimate. The most frequent personality type INFP occurs in 21% of all cases, giving a prediction baseline of 21% to beat.

There are 3 times more introverts than extraverts on the platform. This matches the notion of introverts sharing their thoughts more often online than face-to-face [5] though extraverts might provide the longer texts or the more frequent communication [25]. Sensing people also seem to be rare, of all 8 sensing personality types, only 2 types exceed the threshold for the more robust model. Hence, this adds to critical literature regarding this the sensing/intuition dimension [2].

Though the posts rarely contain “me as an INFP [or any other type]”, the posters tend to stay within their own communities of personality types. This leads to statements such as the following:

“I think the easiest and most efficient approach is a tarp, jigsaw, and mulcher. But that’s just my personal preference. Not all ENTJs are the same.” (ENTJ)

“Hey @MsBossyPants are you down for a debate on Ayn Rand vs Marx? Maybe we should talk about our poor Fi? Oh I know- let’s try to correlate testing ENTJ with being sociopathic. :laughing:” (ENTJ)

¹ <http://personalitycafe.com/forum/>

² <https://www.kaggle.com/datasnaek/mbti-type>



Fig. 2: Histograms of mentioning any personality type for each of the 16 types.

“The last thing my INFJ friend posted on his facebook before committing suicide the next day. Rest in peace <http://vimeo.com/22842206>” (INFJ)

“a pokemon world an infj society everyone becomes an optimist” (INFJ)

Though the users do not refer to themselves when mentioning the type, they talk about their own personality type much more often than about others, see figure 2. As this might bias the performance of the model, an unbiased version of the model is included as well. Mentioning one’s own personality type and staying within a community of one’s own type will however be taken as an important feature for classifying one’s personality since it also allows the use of networks for personality prediction.

3 Methods

The prediction of personalities from text is multiple classes classification problem that requires a stable model that can handle many parameters as text produces a vast amount of model parameters. Support Vector Machines (SVMS) are known to produce stable classification results based on text because they hardly overfit. Additionally, text is usually linearly separable, allowing SVMs to determine a margin between categories [7].

SVMs are mainly used for supervised classification tasks as present here. For linearly separable data, they do not only separate the data by a line but provide a margin between the data points. The width of the margin is determined by the number of support vectors. These support vectors are the data points that are on the margin or within the margin. Usually, this line is a hyperplane in a high-dimensional space. In this high-dimensional space, the data is then linearly separable. The important mechanism for this is the choice of the kernel function that maps the data to this space.

An SVM minimizes the term with quadratic norm $\frac{1}{2}||\mathbf{w}||_2^2$ and allows for a soft margin for error ξ_i . If there are missclassified samples, the first constraint in equation 1 accounts for the error ξ_i caused by these samples. If there wasn't such error term, the data needed to be linearly separable with no missclassified samples or the first constraint could never be met. The error is kept low by being part of the objective function in return.

$$\begin{aligned} \min \quad & \frac{1}{2}||\mathbf{w}||_2^2 + C \sum_{i=1}^m \xi_i \\ \text{s.t.} \quad & y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i \\ & 1 \leq i \leq m \end{aligned} \quad (1)$$

This optimization problem is usually solved using dual formulation with Lagrange multipliers and the KKT, making use of the fact that w can be expressed in terms of the Langrangian multipliers α and y .

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C \\ & \sum_{i=1}^m \alpha_i y_i = 0 \end{aligned} \quad (2)$$

This yields the classification rule similar to the first constraint in equation 1:

$$f(\mathbf{x}) = \text{sgn}(\langle \mathbf{w}, \mathbf{x} \rangle + b) = \text{sgn} \left(\sum_{i=1}^m \alpha_i y_i \langle \mathbf{x}_i, \mathbf{x} \rangle + b \right) \quad (3)$$

The data points that form the margin in the end are the ones where the Lagrangian multipliers are not 0, $\alpha_i \neq 0$.

Pre-processing the data All texts are pre-processed and cleaned, in particular: A DocumentTermMatrix of ngrams (3-grams) was created. The SVM ran with n-grams to preserve all sequential information from the posts. After that, stopwords that carry little or no meaning have been removed. All text was then transformed into lowerspace to account for spelling errors. All word or ngram occurrences have been weighted using tf.idf to even out different document lengths:

$$tfidf_t = f_{t,d} \cdot \log \frac{N}{d_f}$$

$$\begin{aligned} tfidf_t &= \text{weight of term } t \\ f_{t,d} &= \text{occurrences of term } t \text{ in document } d \\ N &= \text{total documents} \\ d_f &= \text{number of documents containing } t \end{aligned} \tag{4}$$

Features General quantitative features such as average word length, total words per person and average words per post have been reported.

Additionally, the relationship between text and personality allows for linguistic features to be manually extracted.

A full set of possibly indicative groups of words and concepts has been added using the word-emotion lexicon by Mohammed & Turney [17]. Especially the intuition/sensing dimension can be distinguished by the use of emotions [4]. Additionally, extraverts post more links [1]. They tend to refer more to themselves and use actual self-representation [25]. Hence, words like "me", "myself" or "I" resemble the feature "self-representation". Extraverts also communicate more and might thereby use more words in general [25], to explain just a few features used.

Following the observations of the data, there will be 6 main models: One model only uses the features, one SVM uses only the text, one then resembles the two. The *text + features* model is then cleaned by removing one's own personality type and predicting only with the 7 more frequent classes.

Hyperparameters were tuned using grid search and a linear kernel. All SVMs were implemented in R using the e1071 library.

4 Results

All SVMs yielded considerably higher results than the baseline. The *text + features* model was able to classify 62% of people's personality types in the test set correctly, the robust and unbiased model 54%, see figure 5. This result including a well-shaped ROC curve, figure 3, answers that at least for this dataset the MBTI can be predicted (1). The confusion matrix shows how the prior distribution of personality types influenced the result: most classification errors happened in the more frequent categories among the over-represented introverts. Interestingly, especially the personality types that differ only in one dimension are more often missclassified. This could imply a general validity of

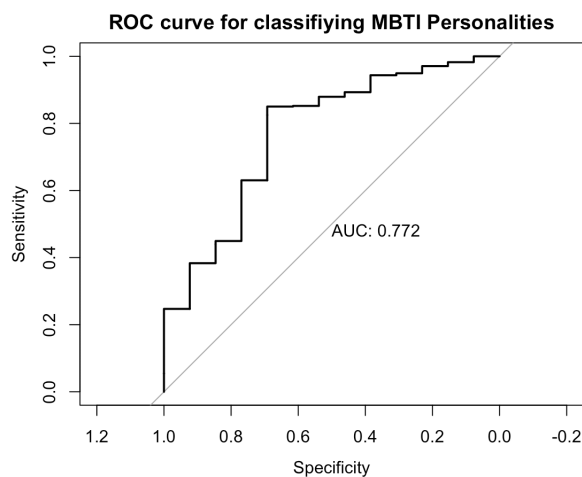
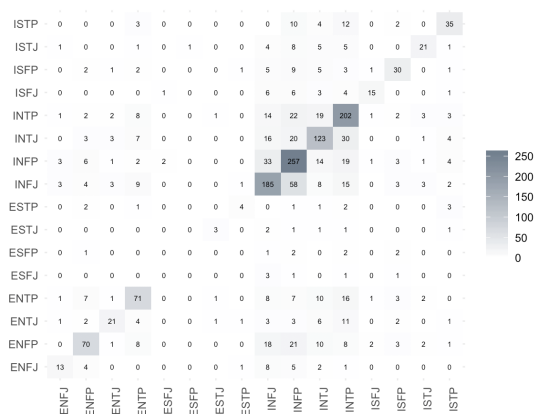


Fig. 3: Aggregated ROC curve for the multi-class SVM.

the MBTI (2): types that don't have much resemblance in theory are usually not confused here either. The *plain features* model comes to 21%. If only the features were used for prediction, they would have to be improved to perform well. As however the personality types are predicted from text, the features may still add some points in accuracy to the model: The text by itself comes to 56%, with features it is 62% accurate.

This result is unfortunately biased. As predicted, mentioning one's own personality even without referring to oneself has a big impact on the model's performance: if one's own personality type is removed from text, it results in



Confusion Matrix

Model	
plain features	0.21
plain text	0.56
text + features	0.62
<i>text + features +</i>	
unbiased w/o type	0.46
robust 7 classes	0.54
robust + unbiased	0.54

Prediction accuracy

Fig. 5: Prediction results

a dropped accuracy of 46%. If however it is altered to a more robust model with the stable classes, it comes to a promising 54%. Overall, this supports the notion of the MBTI being predictable here and indicates proof for research question (1).

5 Discussion

This paper intended to (1) predict MBTI types based on social network posts as done with the Big 5 to then (2) determine if the MBTI is a valid measure for text data. The results show that text can indeed be a good predictor for people’s MBTI personality types. The analysis also showed that similar personality types tend to share communities and hang out with each other. This network effect can be used for future research and implies that network analysis of social media profiles might also predict personality types well.

Though the MBTI is not as accepted as the Big 5 Model, it enjoys a high social validity and a broad acceptance in the population. As many people know their MBTI type and are willing to share it online, they grant access to their personal data including their personality type and allow for easily collectible datasets. With the results of this MBTI analysis, one could then derive 4 of the 5 Big Five dimensions as they correlate, giving estimates for a more valid model and easing out the disadvantage of the MBTI.

Even if researchers have an interest in these large datasets, the implication that personality can be predicted based on public data such as social media posts is a major threat to a user’s data privacy. If the MBTI manages to predict the Big 5, the call for a privacy threat is equally valid for the MBTI. If then the use of data on the Big 5 is heavily restricted [11], one could consider doing the same for MBTI data.

Such private data has been abused in the past, especially concerning targeted ads during the US-American elections by Cambridge Analytica[13]³. The specific manipulation of user’s voting preferences raised the question if targeted ads and thus predicting personality is ethical in the first place.

Participants seem to be willing to share their data with researchers, e.g. in the most prominent example over 30% of myPersonality participants voluntarily shared the contents of their Facebook profiles, together with their personality, intelligence, and other psychometric scores [26]. However, one can hardly speak of *informed consent*[3], especially because the users will not be able to estimate what could be done with the data. This is why Kosinski et al. propose a consent form with clear intentions on how the data will be used [11]. This might not solve the overall issue but could be part of a solution to this ethical problem.

Limitations Not all dimensions delivered equally valid results. Due to the lack of data for the intuition/sensing dimension, a reliable estimate for the

³ “Cambridge Analytica - The Power of Big Data and Psychographics”, <https://www.youtube.com/watch?v=n8Dd5aVXLCc>, retrieved March 14, 2019.

correlated agreeableness dimension in the Big 5 cannot be provided using this dataset.

Though previous researchers reported lower performances around 0.4 or 0.5 [4, 12, 10] and this model seems to perform fairly well, the fact that the dataset comes from a personality forum cannot be discarded. The personality enthusiasts in that forum will probably be much more aware of their type and identify themselves more with it. This would in return make the type much better identifiable for an algorithm in their expressions and their posts. Thus, the results in this analysis will probably be higher than in a more randomly sampled dataset. In future research, a less specific dataset can shed light on the reliability of predicting the Big 5 using the MBTI and thus on the following privacy issues.

References

1. Blackwell, D., Leaman, C., Trampusch, R., Osborne, C., Liss, M.: Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction. *Personality and Individual Differences* **116**, 69–72 (2017). DOI 10.1016/j.paid.2017.04.039. URL <https://linkinghub.elsevier.com/retrieve/pii/S0191886917302891>
2. Boyle, G.J.: Myers-Briggs Type Indicator (MBTI): Some Psychometric Limitations. *Australian Psychologist* **30**(1), 71–74 (1995). DOI 10.1111/j.1742-9544.1995.tb01750.x. URL <http://doi.wiley.com/10.1111/j.1742-9544.1995.tb01750.x>
3. Fleming, D.A., Edison, K.E., Pak, H.: Telehealth Ethics. *Telemedicine and e-Health* **15**(8), 797–803 (2009). DOI 10.1089/tmj.2009.0035. URL <https://www.liebertpub.com/doi/10.1089/tmj.2009.0035>
4. Gjurović, M., Šnajder, J.: Reddit: A Gold Mine for Personality Prediction. In: *Proceedings of the Second Workshop on Computational Modeling of Peoples Opinions, Personality, and Emotions in Social Media*, pp. 87–97. Association for Computational Linguistics, New Orleans, Louisiana, USA (2018). DOI 10.18653/v1/W18-1112. URL <http://aclweb.org/anthology/W18-1112>
5. Goby, V.P.: Personality and Online/Offline Choices: MBTI Profiles and Favored Communication Modes in a Singapore Study. *CyberPsychology & Behavior* **9**(1), 5–13 (2006). DOI 10.1089/cpb.2006.9.5
6. Golbeck, J., Robles, C., Turner, K.: Predicting personality with social media. In: *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11*, p. 253. ACM Press, Vancouver, BC, Canada (2011). DOI 10.1145/1979742.1979614. URL <http://portal.acm.org/citation.cfm?doid=1979742.1979614>
7. Joachims, T.: Text categorization with support vector machines: Learning with many relevant features. In: C. Nédellec, C. Rouveirol (eds.) *Machine Learning: ECML-98*, vol. 1398, pp. 137–142. Springer Berlin Heidelberg. DOI 10.1007/BFb0026683. URL <http://link.springer.com/10.1007/BFb0026683>
8. Jung, C.G., Adler, G., Hull, R.F.C.: *Personality Types*. In: *Collected Works of C.G. Jung*, vol. 6. Princeton University Press, Princeton (2014). URL <http://public.eblib.com/choice/publicfullrecord.aspx?p=1573483>. OCLC: 870244314
9. Kelly, E.L., Conley, J.J.: Personality and compatibility: A prospective analysis of marital stability and marital satisfaction. *Journal of Personality and Social Psychology* **52**(1), 27–40 (1987). DOI 10.1037/0022-3514.52.1.27. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.52.1.27>
10. Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., Graepel, T.: Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning* **95**(3), 357–380 (2014). DOI 10.1007/s10994-013-5415-y. URL <http://link.springer.com/10.1007/s10994-013-5415-y>

11. Kosinski, M., Matz, S.C., Gosling, S.D., Popov, V., Stillwell, D.: Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist* **70**(6), 543–556 (2015). DOI 10.1037/a0039210. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0039210>
12. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* **110**(15), 5802–5805 (2013). DOI 10.1073/pnas.1218772110. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1218772110>
13. Krogerus, H.G.u.M.: Ich habe nur gezeigt, dass es die Bombe gibt. *Tages-Anzeiger* (2018). URL <https://www.tagesanzeiger.ch/ausland/europa/diese-firma-weiss-was-sie-denken/story/17474918>
14. Mairesse, F., Walker, M.A., Mehl, M.R., Moore, R.K.: Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research* **30**, 457–500 (2007). DOI 10.1613/jair.2349. URL <https://jair.org/index.php/jair/article/view/10520>
15. McCrae, R.R., Costa, P.T.: Reinterpreting the Myers-Briggs Type Indicator From the Perspective of the Five-Factor Model of Personality. *Journal of Personality* **57**(1), 17–40 (1989). DOI 10.1111/j.1467-6494.1989.tb00759.x. URL <http://doi.wiley.com/10.1111/j.1467-6494.1989.tb00759.x>
16. McCrae, R.R., John, O.P.: An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality* **60**(2), 175–215 (1992). DOI 10.1111/j.1467-6494.1992.tb00970.x. URL <http://doi.wiley.com/10.1111/j.1467-6494.1992.tb00970.x>
17. Mohammad, S., Turney, P.: Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. In: *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pp. 26–34. Association for Computational Linguistics (2010). URL <http://aclweb.org/anthology/W10-0204>. Event-place: Los Angeles, CA
18. Myers, I.B. (ed.): *MBTI manual: a guide to the development and use of the Myers-Briggs Type Indicator*, 3rd ed edn. Consulting Psychologists Press, Palo Alto, Calif (1998)
19. Oberlander, J., Gill, A.J.: Language With Character: A Stratified Corpus Comparison of Individual Differences in E-Mail Communication. *Discourse Processes* **42**(3), 239–270 (2006). DOI 10.1207/s15326950dp4203_1. URL http://www.tandfonline.com/doi/abs/10.1207/s15326950dp4203_1
20. Ozer, D.J., Benet-Martinez, V.: Personality and the Prediction of Consequential Outcomes. *Annual Review of Psychology* **57**(1), 401–421 (2006). DOI 10.1146/annurev.psych.57.102904.190127. URL <http://www.annualreviews.org/doi/10.1146/annurev.psych.57.102904.190127>
21. Pennebaker, J.W., King, L.A.: Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology* **77**(6), 1296–1312 (1999). DOI 10.1037/0022-3514.77.6.1296. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.77.6.1296>
22. Plank, B., Hovy, D.: Personality Traits on Twitter - or - How to Get 1,500 Personality Tests in a Week. In: *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 92–98. Association for Computational Linguistics, Lisboa, Portugal (2015). DOI 10.18653/v1/W15-2913. URL <http://aclweb.org/anthology/W15-2913>
23. Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our Twitter Profiles, Our Selves: Predicting Personality with Twitter. In: *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, pp. 180–185. IEEE, Boston, MA, USA (2011). DOI 10.1109/PASSAT/SocialCom.2011.26. URL <http://ieeexplore.ieee.org/document/6113111/>
24. Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M.E.P., Ungar, L.H.: Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLoS ONE* **8**(9), e73791 (2013). DOI 10.1371/journal.pone.0073791. URL <https://dx.plos.org/10.1371/journal.pone.0073791>

25. Seidman, G.: Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and Individual Differences* **54**(3), 402–407 (2013). DOI 10.1016/j.paid.2012.10.009. URL <https://linkinghub.elsevier.com/retrieve/pii/S0191886912004916>
26. Stillwell, D.J., Kosinski, M.: myPersonality project Web site (2018). URL <http://mypersonality.org>
27. Tett, R.P., Jackson, D.N., Rothstein, M.: Personality Measures As Predictors of Job Performamnce: A Meta-Analytic Review. *Personnel Psychology* **44**(4), 703–742 (2006). DOI 10.1111/j.1744-6570.1991.tb00696.x. URL <http://doi.wiley.com/10.1111/j.1744-6570.1991.tb00696.x>
28. Youyou, W., Kosinski, M., Stillwell, D.: Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences* **112**(4), 1036–1040 (2015). DOI 10.1073/pnas.1418680112. URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1418680112>