

## Article

# Recreating the Relationship between Subjective Wellbeing and Personality Using Machine Learning: An Investigation into Facebook Online Behaviours

Alexandra Marinucci \*, Jake Kraska and Shane Costello 

Faculty of Education, Monash University, Clayton 3168, VIC, Australia; jake.kraska@monash.edu (J.K.); shane.costello@monash.edu (S.C.)

\* Correspondence: ally.marinucci94@gmail.com; Tel.: +61-3-9905-0283

Received: 29 June 2018; Accepted: 3 September 2018; Published: 5 September 2018



**Abstract:** The twenty-first century has delivered technological advances that allow researchers to utilise social media to predict personal traits and psychological constructs. This article aims to further our understanding of the relationship between subjective wellbeing (SWB) and the Five Factor Model (FFM) of personality by attempting to replicate the relationship using machine learning prediction models. Data from the myPersonality Project was used; with observed SWB scores derived from the Satisfaction With Life Scale (SWLS) and Five Factor Model (FFM) personality profiles generated using responses on the 100-item IPIP proxy of the NEO-PI-R. After data cleaning, FFM personality traits and SWB scores were predicted by reducing Facebook Likes into 50 dimensions using SVD and then running the data through six multiple regressions (fitting the model via least squares and splitting the data via k-folds validation) with the Likes dimensions as predictors and each of the FFM traits and the SWB score as response variables. Standard multiple regression analyses were conducted for the observed and machine learning predicted variables to compare the relationships in the context of previous literature. The results revealed that in the observed model, high SWB was predicted by high extraversion, conscientiousness, and agreeableness, and low openness to experience and neuroticism as per previous research. For the machine learning model, high SWB was predicted by high extraversion, openness to experience, conscientiousness, and agreeableness, and low neuroticism. The relationships between SWB and extraversion, neuroticism, and conscientiousness were successfully replicated in the machine learning model. Openness to experience changed direction in its relationship with SWB from the observed to machine learning-derived variables due to failure to accurately recreate the variable, and agreeableness was multicollinear with SWB in the machine learning model due to the unknowing use of identical digital behaviours to replicate each construct. Implications of the results and directions for future research are discussed.

**Keywords:** machine learning; personality; wellbeing; myPersonality project; Five Factor Model; online behaviour; Facebook; Likes; social media

## 1. Introduction

Fast-paced technological trends demand research tools in psychology to evolve. There has been a historical focus on self-report methods and traditional behaviour analysis due to their ease of use and proliferation in psychological research. Novel approaches such as machine learning and data mining have recently begun to gain traction in psychological research [1]. Data mining allows large, diverse samples to be analysed and utilised in algorithms to predict future outcomes [2]. In particular, the algorithms can be used to predict psychological constructs such as subjective wellbeing (SWB) and the traits of the Five Factor Model (FFM) of personality [3]. Previous research has demonstrated

machine learning's ability to recreate psychological constructs from online data. However, the observed relationship between wellbeing and personality has not yet been recreated via machine learning techniques in the extant literature. The current study utilises simple machine learning techniques such as singular value decomposition, k-folds validation, and linear regression. Facebook 'Like' data was first reduced via unsupervised feature extraction using singular value decomposition, employing the dimensions and pre-labelled personality (FFM) and SWB data, and linear regression and k-folds validation were used to predict participants' personality (FFM) and SWB. These predicted values were then used to recreate the relationship between the FFM and SWB, and assess the accuracy of the prediction model compared to observed scores.

### 1.1. Subjective Wellbeing

Research surrounding SWB has captivated the field of psychology. Brickman and Campbell [4] coined the term of the 'hedonic treadmill', in which individuals change in reaction to improved circumstances, such as wealth and material goods, yet do not yield happiness. They and other researchers found that an increase in income did not increase one's SWB and found that lottery winners were typically less happy and paraplegics happier than one would anticipate [5–7]. On the other hand, the most significant influences of SWB have been found to be personality traits, as they predispose an individual to life experiences and behaviours that may positively or negatively affect one's average level of life satisfaction [8–10]. A higher level of SWB would be associated with frequent positive, pleasant affective experiences. An individual may consider this in the form of cognitions, such as evaluation of marital and career satisfaction, or in the form of affect, such as experiencing certain moods and emotions in reaction to an event [11]. Investigating SWB and how it is best predicted is important in psychology to further our understanding of mental illnesses, such as depression, and researchers' and society's perception of true 'happiness'.

### 1.2. The Five Factor Model Model of Personality

An individual's personality may be described in layman's terms as 'friendly', 'outgoing', 'loud', or 'shy'. However, the dominant understanding of personality in the academic community is the FFM, constituted of: extraversion, neuroticism, openness to experience, conscientiousness, and agreeableness. Personality psychology aims to understand the whole person to comprehend how individuality is organised and integrated [12,13]. The FFM traits have been well established in psychological research as a way to address the underlying variety in human behaviour with a nomenclature that can classify individual personality differences [14]. Additionally, personality has been linked to marital and relationship outcomes, career satisfaction, social adaptation, and cultural differences [15–17].

### 1.3. Machine Learning

Machine learning is a relatively new and emerging research tool in psychological study. Due to the evolving nature and skills required in this field, computer scientists and engineers have typically dominated; however, it is becoming a popular tool in social sciences [2,18]. Machine learning involves applying a performance algorithm to a large data set to produce a prediction model and using this model to predict an outcome [19]. Repeating this process iteratively allows for a 'perfected' model and accurate predictions of psychological constructs. In order to use machine learning, data mining is required, in which large data sets can be utilised to create new information or strengthen previous knowledge. Feldman, et al. [20] used Facebook 'status updates' of 73,789 participants recruited through the myPersonality application to determine whether the machine learning condition supported the previously found positive relationship between profanity and honesty. Utilising linguistic inquiry and word count on the large data set replicated this previous relationship, showing that those who used more profanity in their Facebook 'status updates' were more likely to be honest. Data mining also allows for the analysis of large data sets with a higher accuracy and thus additional insights can be determined [21].

Two types of machine learning are commonly used in psychology, which vary through the labelling of data. Supervised machine learning requires prior labelling of the data by the researcher based on theory or previous knowledge, and produces an algorithm that can be used to predict future instances. On the other hand, unsupervised machine learning does not involve prior labelling of the data and instead, the machine applies its own clustering of data to produce an unknown class of items and predictions [22].

#### *1.4. Advantages and Disadvantages of Machine Learning*

Machine learning brings advantages in terms of large sample sizes, generalisation of research, a considerably lower cost, a higher statistical power, and lower bias in results. Stillwell and Kosinski [23], among other researchers [1,2], have commented on the usefulness of internet-based psychological research and its ability to yield large samples of participants, allowing a larger scale to base research on. Similarly, Kosinski, et al. [24] noted the proficiency of Facebook as a research tool, permitting large and non-WEIRD (western, educated, industrialised, rich, democratic) samples to be inexpensively recruited. They also stated that Facebook provides records of human behaviour expressed in a natural online environment. This reduces reference group effect bias, typically found in self-report questionnaire measures (such as the IPIP; International Personality Item Pool), which refers to describing oneself in relation to others. Human activity monitored by digital services and devices allows behaviours to be digitally mediated, permitting large-scale samples to be obtained with the minimisation of sampling error and reduction of group effect bias [24,25]. Using the Internet as a research tool permits the collection of diverse samples and generally produces better quality data [1].

Although machine learning brings great advantages for research, disadvantages of this tool still exist. Human perception is flexible and can recognise behavioural cues that are not matched by computer-based predictions. Assessing and determining personality not only depends on questionnaire outcomes, but also on how the individual behaves subconsciously and certain cues that may pertain to dishonesty [26]. For Facebook data, users are able to remove online behavioural traces, which may render their profile subject to misinterpretation and social desirability bias when analysed via machine learning techniques [24]. Finally, social media platforms are vulnerable to fake profiles, which could skew and affect research results. However, Kosinski, Matz, Gosling, Popov and Stillwell [24] have argued that these profiles are usually easy to detect. Because these disadvantages can be controlled for, machine learning poses a promising future for psychological research.

#### *1.5. Social Media Data in Psychology Research*

Researchers have investigated how Facebook data can be used to produce algorithms to predict psychological constructs. Website choice, 'Like'-based, and language-based data have been the most commonly used variables [26,27]. For example, most users on Twitter were classified as emotionally stable and extroverted by using counts of the Twitter information 'following', 'followers', and 'listed' [28]. Additionally, Reece and Danforth [29] successfully identified markers of depression from participants' Instagram photos, which surpassed general practitioners' typical unassisted diagnostic success rate for depression. Quercia, et al. [30] evaluated the efficacy of digital methods to predict the strength of online social relationships on Facebook and their findings were consistent with previous analyses used on Twitter [31]. The researchers concluded that explanations for relationship deterioration did not differ between online and offline social worlds. Kosinski, Stillwell and Graepel [3] praised the high predictive power of Facebook 'Likes' and predicted a range of personal attributes with varying accuracy. A large sample of Facebook information for approximately 58,000 participants allowed for the development of a prediction model in which numeric variables such as age and intelligence were predicted using linear regression. Dichotomous variables such as gender were predicted using logistic regression. Ethnic origin, gender, and age were the most accurately predicted variables. Personality traits were moderately predicted, whereas SWB was weakly predicted. The researchers attributed the low accuracy of prediction for SWB to the basis that Facebook 'likes'

accrue over a continuous period and give a long-term score of SWB, which is not reflected in the SWLS (Satisfaction with Life Scale) because that is a snapshot in time. Social media and technology are an important part of society and integrating them into psychological research allows human behaviour to be monitored and analysed in a way that may be beneficial to future research [32].

### 1.6. The Relationship between Subjective Wellbeing and the Five Factor Model Model of Personality

Previous findings indicate that there is a strong relationship between the FFM of personality and SWB. High levels of SWB are associated with low levels of neuroticism and high levels of conscientiousness, extraversion, agreeableness, and openness to experience [8–10,33,34]. Fujita [35] found a strong correlation between neuroticism and negative effect, whilst a meta-analysis determined a moderate correlation between extraversion and positive effect [36]. The consistency of these correlations displays the effectiveness of predicting SWB from the FFM of personality. Correlational studies suggest that individuals are sensitive to certain stimuli and thus will respond to events differently [37–39]. Personality type predisposes an individual to experiencing certain life events, which in turn affects an individual's level of SWB [40].

### 1.7. The Current Study

Studies have shown the ability to predict a person's personality at a superficial level using machine learning [2,3,23,25,26,28,41]; however, the relationship between machine learning predicted SWB and personality has not yet been explored. Kosinski, Stillwell and Graepel [3] found the test retest reliability of the openness to experience FFM trait to approximate the predicted versus observed correlation score, suggesting that the observation of user Facebook 'Likes' is about as informative as that from a personality questionnaire test score. The research question addressed in the current study is whether machine learning and data mining can be used to predict personality (FFM) and SWB, and then whether these outcomes can be used to replicate previously observed relationships between the FFM of personality and SWB. It is predicted that the machine learning model will accurately predict SWB and FFM, resulting in a consistent relationship between observed FFM traits and observed SWB and that of predicted FFM traits and predicted SWB.

## 2. Materials and Methods

### 2.1. Participants

The data was obtained from the "myPersonality Project" (mypersonality.org; [23]). The myPersonality Project contains more than four million individual Facebook profiles. The participants had accessed the myPersonality application through their Facebook profiles during the years 2007 to 2012. The current analysis was conducted in two steps, substantially reducing the number of participants (see [bitbucket.org/jakekraska/swlbig5](http://bitbucket.org/jakekraska/swlbig5) for data reduction and analysis code).

Step one involved identifying participants that had provided demographic, SWB, and FFM data and merging this together, resulting in 80,628 participants. Participants missing from the myPersonality user-likes file were removed, resulting in a dataset that included 26,573 participants that had complete FFM personality scores and a SWB score. After removing missing data and duplicate data, iteratively removing participants that had less than 10 likes and then removing likes that have been liked less than 50 times, 21,122 participants and 10,377 Likes remained. The singular value decomposition and least squares multiple linear regression analyses were conducted for this data, predicting FFM personality scores and SWB scores from 50 SVD dimensions across 10 folds (predicting values for 10% of the sample from 90% of the sample, iteratively). This format of validation (k-folds validation with 10 folds) was utilised due to the small number of participants available after data cleaning. Available country data (148 countries) for these participants is included in Appendix A (Table A1).

For the second stage, remaining analyses including correlations between the original and predicted values, and a comparison of the multiple linear regression models, were only conducted for those

participants that met the inclusion criteria. That is, only participants that were aged between 16 and 90, and those that had provided a gender were included in the final analysis. The average age of participants in the final sample ( $n = 13,497$ ) was 24.56 years ( $SD = 7.08$ ), consisting of less male participants ( $n = 5322$ , 39.43%) than females ( $n = 8175$ , 60.57%). Participants who were aged outside the range of 17 to 90 were omitted due to ethical guideline considerations and false ages given (e.g., 150 years). Due to the anonymity of the data, ethics exemption was granted by the Monash University Human Research Ethics Committee.

## 2.2. Subjective Wellbeing Measure

To measure subjective wellbeing (SWB), the Satisfaction with Life Scale (SWLS) from Diener, et al. [42] was administered via the myPersonality application. The five-item SWLS is a widely used and reliable measure for SWB. A review confirmed the ability of the SWLS to measure SWB as a cognitive judgemental process with a high internal consistency and temporal reliability [43]. Using a Likert scale (ranging from 1 = strong disagree to 7 = strongly agree), participants are asked to respond to five questions about how they view their own life. For example: *'In most ways my life is close to my ideal'*. A low overall score indicates extreme dissatisfaction with life and a high score indicates extreme satisfaction with life.

An internal consistency coefficient of 0.87 and a test-retest correlation coefficient of 0.82 over a two month period was reported by Diener, Emmons, Larsen and Griffin [42]. A later study found a moderate mean Cronbach's alpha coefficient ( $\alpha = 0.78$ ) for the SWLS and attributed the moderate score to the small number of items in the scale [44].

## 2.3. Five Factor Model Personality Measure

The independent variables consisted of the five personality traits from the FFM and were measured using the 100-item IPIP proxy of the NEO-PI-R through the myPersonality application on Facebook [24]. The NEO-PI-R is a widely used and comprehensive measure of an individual's FFM of personality—extraversion (ext), neuroticism (neu), openness to experience (ope), conscientiousness (con), and agreeableness (agr) [45,46]. The NEO-PI-R demonstrates a high internal consistency and stability over a six-year period, showing the reliability and validity of the measure [45,47,48]. The five subscales in the 100-item IPIP proxy have 20 items for which the participants must respond on a five-point Likert scale. For example: responding strongly agree to *'I know how to captivate people'* would contribute to a higher score on the extraversion scale.

The reliability coefficients for the 100-item IPIP proxy of the NEO-PI-R range from 0.85 for the agreeableness scale to 0.91 for both the neuroticism and extraversion scales [49].

## 2.4. Data Analysis

Data analysis was conducted using R version 3.5.1 [50] and RStudio, version 1.1.453 [51]. The methodology is contained in Figure 1.

The statistical procedure was modelled on 'Mining Big Data to Extract Patterns and Predict Real-Life Outcomes' by Kosinski, Wang, Lakkaraju and Leskovec [2], as well as other research and guidelines in the area of investigation, e.g., [26,52]. Six data sets were utilised that each had specific and not necessarily the same users: a final SWB score, final scores of each FFM trait, Facebook likes of the user, a list of Facebook like ids and their names, demographic (age, gender) data, and location (country).

For the prediction of each variable, a user-like matrix was constructed to match users from the SWLS and FFM data sets and likes. The matrix was trimmed, removing users with less than 10 likes and like entities (e.g., "Sleeping Too Much", "Saying I love you", "Jason Mraz", "Bowling", "Talk With a British Accent Day!") with less than 50 users. The remaining users were split into 10 folds to reduce overfitting. Fifty SVD dimensions were extracted and underwent Varimax rotation. Figure 2 displays the scree plot for the SVD of the user-like matrix.



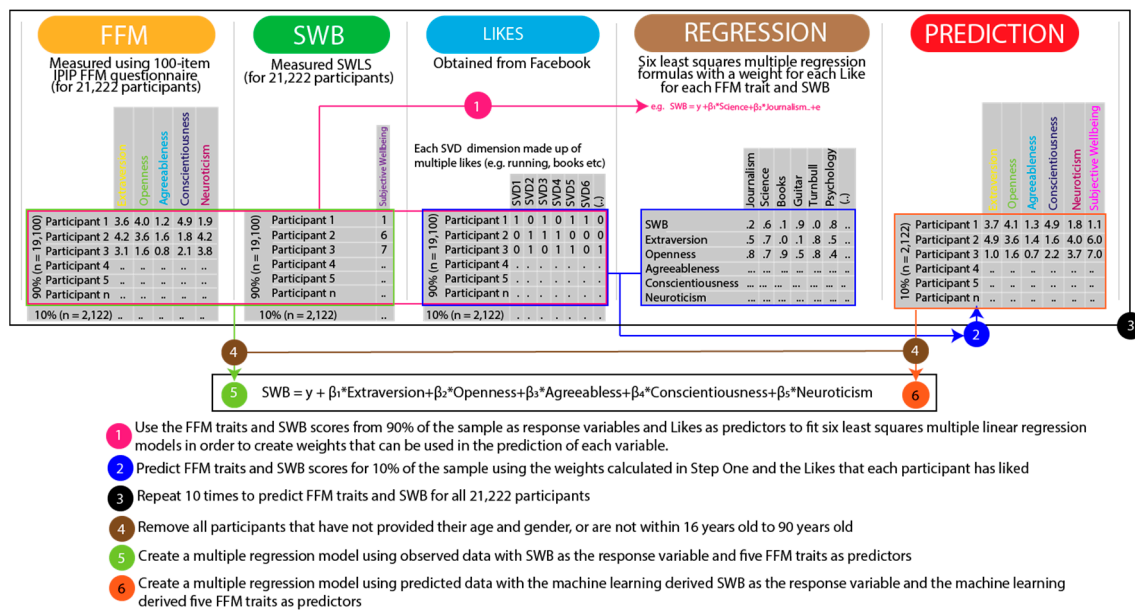


Figure 1. A graphical representation of the research design.

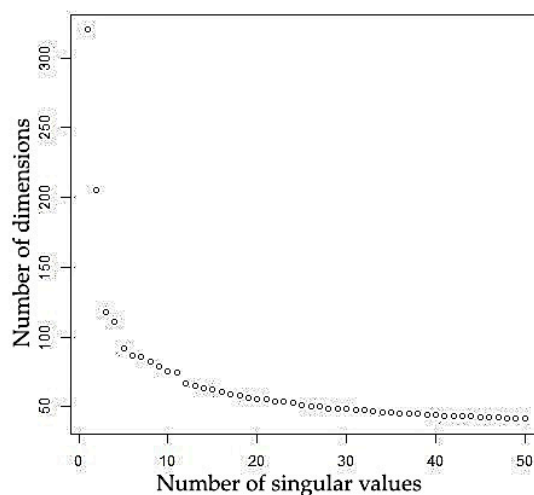


Figure 2. Scree plot for SVD of the user-like matrix.

Multiple linear regression (least squares used for fitting the model), SVD, and k-folds validation ( $k = 10$ ) were used to predict FFM personality traits and SWB scores for participants ( $n = 21,122$ ). Participants that did not include gender and the age range of 16–90 years old were then removed. A correlation analysis was performed for each predicted and observed variable to determine whether it had been replicated accurately. A multiple linear regression model ( $n = 13,497$ ) was built for the observed data (observed SWB as the response variable and each observed FFM personality trait as the predictor variables) and the predicted data (predicted SWB as the response variable and each predicted FFM personality trait as the predictor variables). Correlation, ANOVA, and covariance analyses were run to determine whether the relationship between SWB and personality had been replicated.

### 3. Results

Descriptive statistics for the observed variables prior to age and gender matching are shown in Table 1. Descriptive statistics for the predicted variables are shown in Table 2. After deriving the predicted scores for each variable through a machine learning algorithm, preliminary analyses were

conducted to ensure no violation of normality and homoscedasticity. With the use of a  $p < 0.001$  criterion, Mahalanobis distance and Cook's distance did not suggest the presence of any outliers (Max  $MD(12) = 24.38$ , Max  $D_i$  Observed = 0.002, Max  $D_i$  ML = 0.152). The variance inflation factors for the 12 variables in the regression models were less than 10, indicating the absence of collinearity in the sample. The correlation between agreeableness and SWB in the machine learning regression model suggested multicollinearity ( $r = 0.650$ ). However, as the variance inflation factor did not suggest collinearity and SWB is the dependent variable, both were retained. Correlations of the observed variables are contained in Appendix C Table A4 and for the predicted variables, in Appendix C Table A5.

**Table 1.** Descriptive statistics for the observed FFM traits and SWB ( $n = 21,122$ ).

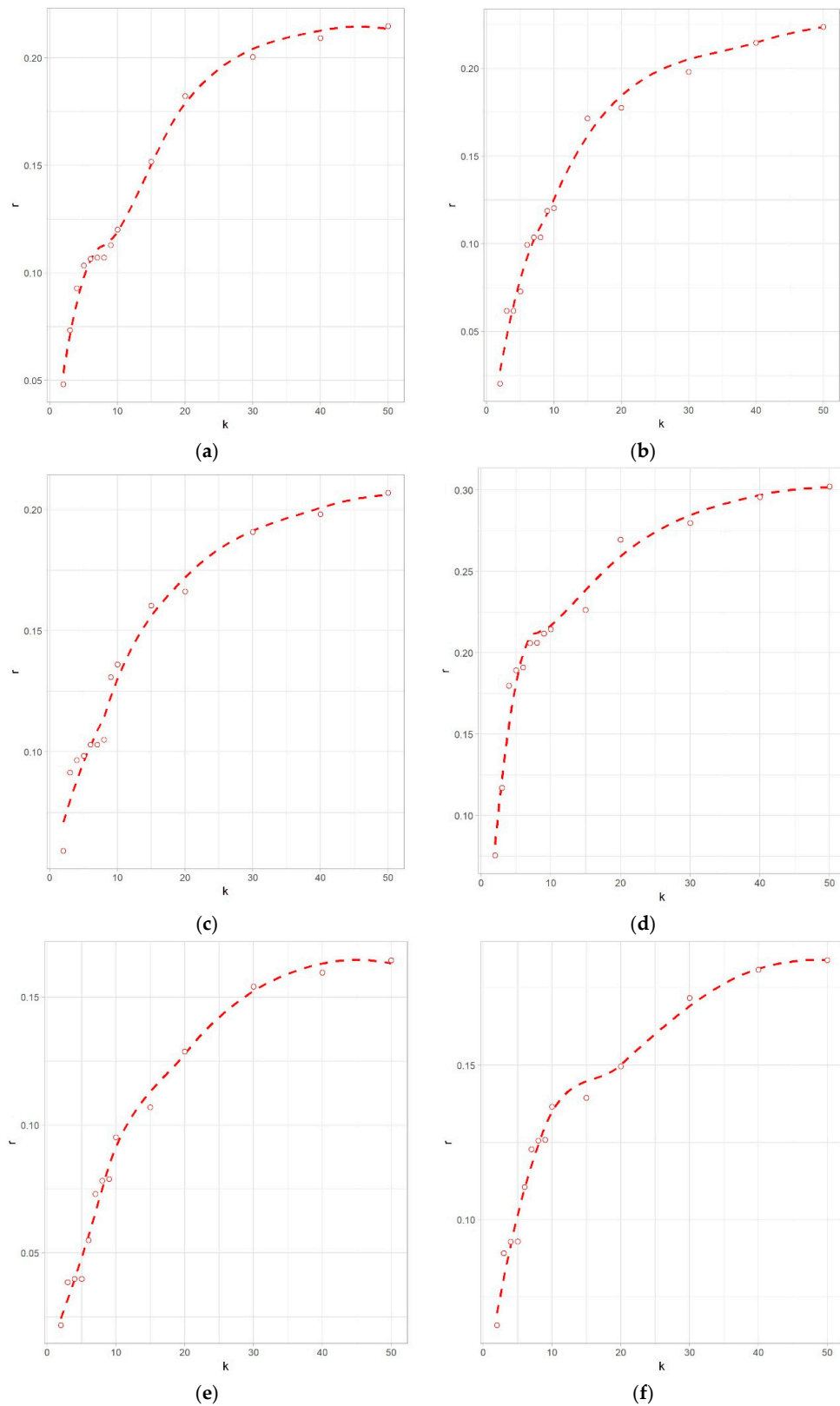
	<i>M</i>	<i>Md</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>	<i>Range</i>	<i>Skew</i>	<i>Kurtosis</i>	<i>Std. Error</i>
SWB	4.34	4.40	1.37	1.00	7.00	6.00	−0.27	−0.81	0.01
Ext	3.32	3.35	0.83	1.00	5.00	4.00	−0.20	−0.57	0.01
Ope	4.08	4.15	0.56	1.10	5.00	3.90	−0.76	0.67	0
Agr	3.51	3.55	0.65	1.00	5.00	4.00	−0.43	0.01	0
Con	3.33	3.33	0.71	1.00	5.00	4.00	−0.07	−0.35	0
Neu	2.88	2.85	0.84	1.00	5.00	4.00	0.10	−0.57	0.01

**Table 2.** Descriptive statistics for the predicted FFM traits and SWB score ( $n = 21,122$ ).

	<i>M</i>	<i>Md</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>	<i>Range</i>	<i>Skew</i>	<i>Kurtosis</i>	<i>Std. Error</i>
SWB	4.34	4.37	0.22	2.41	5.78	3.37	−1.16	7.72	0
Ext	3.32	3.32	0.17	1.85	4.94	3.09	0.45	8.75	0
Ope	4.08	4.05	0.17	2.71	5.56	2.85	1.45	7.41	0
Agr	3.51	3.51	0.10	2.72	4.43	1.71	0.03	7.42	0
Con	3.33	3.34	0.14	2.09	4.43	2.34	−1.16	8.16	0
Neu	2.88	2.86	0.17	1.64	4.37	2.73	0.74	7.51	0

### 3.1. Singular Value Decomposition Analysis (SVD)

The results from the SVD analysis are presented in Figure 3a–d;  $r$  denotes the prediction accuracy and  $k$  denotes the number of dimensions. According to Gignac and Szodorai [53], the normative guidelines for small, typical, and large effect sizes are  $r = 0.10$ ,  $r = 0.20$ , and  $r = 0.30$ , respectively, which will be followed for this analysis. Overall neuroticism, extraversion, and openness to experience were predicted with the greatest accuracy. All variables appear to be of a typical to large effect size in replicating the observed variables.



**Figure 3.** Accuracy of: (a) Neuroticism across  $k$  dimensions, (b) Extraversion across  $k$  dimensions, (c) Conscientiousness across  $k$  dimensions, (d) Openness to Experience across  $k$  dimensions, (e) Agreeableness across  $k$  dimensions, and (f) SWB across  $k$  dimensions. It is important to note that the accuracy plots utilised predicted data that was obtained from training the algorithm on 100% of the sample, rather than iteratively through 10 folds.



### 3.2. Correlational Analysis (Accuracy of Predictions)

Table 3 displays the correlations between the observed and predicted scores for those participants aged 16–90 years old. Unlike in the SVD analysis above, the predicted data used in this analysis was obtained through training the model on 10 folds, in order to reduce overfitting. According to Gignac and Szodorai [53], there is a moderate relationship between the scores for SWB, extraversion, agreeableness, conscientiousness, and neuroticism, suggesting a relationship between the observed and machine learning-derived scores for these variables. Openness to experience has a large relationship, suggesting a strong relationship between the observed and machine learning-derived scores.

**Table 3.** Correlations between observed and machine learning-derived FFM traits and SWB.

Variable	Correlation
SWB	0.17
Ext	0.21
Ope	0.31
Agr	0.15
Con	0.20
Neu	0.20

### 3.3. Multiple Linear Regression Models for Observed and Predicted Data

The two multiple linear regression model statistics for the models of observed scores for SWB and the FFM personality traits and the machine learning-derived (predicted) scores of the same variables are summarized in Table 4. For the observed scores, a standard multiple regression was performed for SWB and extraversion, openness, agreeableness, conscientiousness, and neuroticism.  $R$  was significantly different from zero,  $F(5, 13,491) = 992.6$ ,  $p < 0.001$ , with adjusted  $R^2$  at 0.269, and therefore approximately a quarter of the variability in SWB is predicted by the FFM personality traits. Extraversion, agreeableness, conscientiousness, and neuroticism significantly predicted SWB. Extraversion uniquely explained 1.07%, openness uniquely explained 0.01%, neuroticism uniquely explained 8.26%, conscientiousness uniquely explained 1.66%, and agreeableness uniquely explained 0.25% of the variance in SWB.

For the machine learning-derived scores, the same process was repeated. Again,  $R$  was significantly different from zero,  $F(5, 13,491) = 3585$ ,  $p < 0.001$ , with adjusted  $R^2$  at 0.570, and therefore more than half of the variability in SWB is predicted by the FFM personality traits when using SWB and FFM variables that were predicted using machine learning techniques. For the derived scores, all independent variables significantly predicted SWB. The machine learning-derived factor extraversion uniquely explained 0.18%, openness uniquely explained 9.96%, neuroticism uniquely explained 0.94%, conscientiousness uniquely explained 2.33%, and agreeableness uniquely explained 17.09% of the variance in SWB. See Appendix B for the covariance matrices produced from the two regression models (see Tables A2 and A3).

**Table 4.** Multiple Linear Regression Model of observed FFM traits on SWB and of machine learning-derived FFM traits on the machine learning-derived SWB ( $n = 13,497$ ) after matching predicted values with gender and age.

IV	B	SE	<i>t</i>	<i>p</i>	<i>sr</i> <sup>2</sup>	VIF	Confidence Interval	
							Lower	Upper
Ext	0.188	0.013	13.957	<0.0001 <sup>2</sup>	0.012	1.213	0.161	0.214
ML <sup>1</sup> Ext	0.159	0.008	20.247	<0.0001 <sup>2</sup>	0.002	1.173	0.144	0.175
Ope	−0.019	0.018	−1.014	0.310	<0.0001	1.038	−0.055	0.017
ML Ope	0.044	0.008	5.615	<0.0001 <sup>2</sup>	0.100	1.150	0.029	0.061
Agr	0.136	0.017	8.066	<0.0001 <sup>2</sup>	0.003	1.175	0.103	0.169
ML Agr	1.022	0.015	69.586	<0.0001 <sup>2</sup>	0.171	1.417	0.993	1.051
Con	0.257	0.015	16.825	<0.0001 <sup>2</sup>	0.017	1.134	0.227	0.286
ML Con	0.202	0.015	16.353	<0.0001 <sup>2</sup>	0.023	1.850	0.177	0.226
Neu	−0.611	0.014	−43.074	<0.0001 <sup>2</sup>	0.083	1.400	−0.639	−0.583
ML Neu	−0.425	0.009	−48.502	<0.0001 <sup>2</sup>	0.009	1.372	−0.443	−0.408

<sup>1</sup> ML = machine learning-derived variable, <sup>2</sup> Significant at: 0.001.

#### 4. Discussion

The current study aimed to replicate the relationship between the FFM of personality and SWB using simple machine learning techniques: singular value decomposition, k-folds validation, and multiple linear regression.

It was hypothesised that the machine learning model would accurately recreate the SWB and FFM variables. The results support this hypothesis to an extent. From the correlation analysis, the observed scores for extraversion ( $r = 0.21$ ), neuroticism ( $r = 0.20$ ), and openness ( $r = 0.31$ ) were most accurately recreated in the machine learning model.

The second hypothesis postulated that the variables predicted through machine learning techniques would be capable of replicating the relationship between observed SWB and the FFM variables. Again, this hypothesis is partially supported. Higher scores for extraversion, agreeableness, conscientiousness, and neuroticism predicted a higher score for SWB; however, the openness to experience prediction reversed in the machine learning model. The openness to experience prediction reversed in direction, so in the machine learning model, an increase in openness to experience predicted a higher SWB. This may be attributed to the failure to recreate the variable in the first hypothesis. Based on the multiple regression analyses, openness to experience became positively correlated with SWB in the machine learning model, which could be attributed to reference group bias through the administration of the NEO-PI-R via Facebook [24,25]. A higher correlation was found between agreeableness and SWB, which is due to the unknowing use of identical digital behaviours to predict the variables or multicollinearity, which inflated the relationship. This suggests that using machine learning to recreate variables is likely to overestimate the relationship between variables.

Whilst the findings from this study pose additional evidence for the utility of using digital behaviour as data to produce prediction models, the accuracy of predictions for the currently investigated constructs is not high when relying solely on SVD and multiple linear regression. Kosinski, Wang, Lakkaraju and Leskovec [2] seem to inflate the accuracy of the linear regression model to predict psychological constructs; while they achieved a high accuracy when predicting gender, personality was predicted with a relatively lower accuracy. Additionally, Kosinski, Stillwell and Graepel [3] found prediction of personality constructs to range from  $r = 0.17$  to  $r = 0.43$ , which are not necessarily high accuracies. The prediction for SWB ( $r = 0.17$ ) was very low in comparison to what was found for age ( $r = 0.75$ ) in their study. Further investigation within the psychological literature into more complex machine learning techniques that may increase the accuracy of predictions using social media data is required.

#### 4.1. The Relationship between SWB and the FFM of Personality

The correlations between the SWB and FFM variables in both models (observed and machine learning-derived) partly replicated previous literature. A summary of correlations between the FFM traits and SWB from four studies in the literature is displayed in Table 5. Extraversion and neuroticism were replicated with a reasonable accuracy and mirrored the findings of Steel, Schmidt and Shultz [8] and Grant, Langan-Fox and Anglim [33]. Openness to experience replicated the findings from Steel, Schmidt and Shultz [8] and Grant, Langan-Fox and Anglim [33] in the original model; however, the machine learning model did not accurately replicate the variable. To an extent, the observed variable of conscientiousness paralleled the findings from Anglim and Grant [34], though the machine learning variable almost doubled in its correlation size and did not represent any previous findings in the literature. Agreeableness in both models did not represent or mirror any previous research.

**Table 5.** Correlations from the literature: FFM Traits with SWB.

	Neu	Ext	Ope	Agr	Con
Observed Scores ( $n = 13,497$ ) <sup>1</sup>	−0.49	0.29	0.05	0.25	0.29
Predicted Scores ( $n = 13,497$ ) <sup>1</sup>	−0.54	0.36	−0.09	0.66	0.54
Steel, Schmidt & Shultz (2008)	−0.38	0.28	0.03	0.14	0.22
DeNeve & Cooper (1998)	−0.24	0.17	0.14	0.16	0.22
Anglim & Grant (2016)	−0.57	0.51	0.13	0.11	0.35
Grant et al. (2009)	−0.36	0.22	0.04	0.16	0.21

<sup>1</sup> Observed scores and Predicted scores = current study correlations, Neu = Neuroticism, Ext = Extraversion, Ope = Openness to Experience, Agr = Agreeableness, Con = Conscientiousness.

None of the variables in either model replicated the findings from DeNeve and Cooper [9], which could be attributed to the meta-analysis' mean age of 53 years. The current study was predominantly young adult aged and therefore the different life stage may explain the discrepancy [54,55]. Only conscientiousness in the original model slightly mirrored Anglim and Grant [34], which could be due to the different personality measure used, the 30-item Facet IPIP. The shorter scale may have exaggerated the scores for neuroticism and extraversion, as they are considerably higher in comparison to the other studies mentioned.

The correlations between the observed and machine learning predicted variables somewhat replicated the findings of Kosinski, Stillwell and Graepel [3]. The highest correlation for the current study is for extraversion and the SWB correlation in the current study was almost the same as that found by the researchers ( $r = 0.17$ ). As Kosinski, Stillwell and Graepel [3] did not specify the correlations for the other FFM variables ( $r = 0.17$  to  $r = 0.30$ ), conclusions regarding these variables are not complete. However, agreeableness, conscientiousness, and neuroticism were in the range stated by the researchers. These similarities are to be expected given that we have used the same initial dataset, but with different inclusion criteria.

Overall, for the original model with observed variable scores, high SWB was predicted by high extraversion, agreeableness, and conscientiousness, and low openness to experience and neuroticism. For the machine learning model, high SWB was predicted by high extraversion, openness to experience, conscientiousness, and agreeableness, and low neuroticism. Therefore, it could be concluded that high extraversion, conscientiousness, and agreeableness, and low neuroticism, are relatively consistent predictors of high SWB.

The greater prediction accuracy of the machine learning model linear regression compared to the observed data linear regression (Table 4) may be due to the genuine nature of Facebook 'likes' used to train the machine learning algorithm, and thus their impact on the recreated variables. When considering the machine learning model to predict the variables, most variables recreated the observed variables with a relative accuracy with a large effect size according to Gignac and Szodorai [53]. Using social media data to predict real life outcomes presents an important opportunity

in psychology to further measure how individuals can be perceived and how they behave in a natural online environment [24].

#### 4.2. Implications, Limitations and Future Research

The basis of Facebook ‘likes’ is to record human behaviour through expressing positive opinions regarding online content. Technological advances have allowed big data to be extracted from social media websites and this data can be manipulated and analysed to further understand human behaviour [32]. The amount of information that can be gathered through social media is significant and generates new areas and possibilities for future research. The current study had a large sample size of 21,112 participants (used to predict FFM traits and SWB for 13,497 participants aged 16–90) from 148 different countries, which exhibits the advantage of large samples allowing for high statistical power to be obtained [2]. Despite previous authors praising the utility of social media to attract a less western population, the sample for this study was predominately western, as most participants were from Australia, Canada, the United Kingdom, or the United States, limiting the generalisability of the results. Future research should investigate non-westernised countries and the prediction models based off their Facebook ‘likes’, as they may be considerably different from the western population.

Although using the Internet and Facebook information in psychological research reduces reference group bias, some bias may be evident. Though the Internet provides a medium to observe human behaviour, individuals can still put on a façade and “fake good”. As of 2017, over two million applications exist (Apple and Android) to alter photos (similar to Adobe’s ‘Photoshop’), access social media sites, locate oneself on a map, order food delivery and transport, track health and exercise, and much more. Holland and Tiggemann [56] systematically reviewed 20 studies and concluded that social networking website use, body image, and disordered eating are related (regardless of gender). Particularly viewing and uploading photos and attention seeking ‘status updates’ that received negative responses were damaging. Another study found, in a sample of adolescent females, that increased appearance exposure on Facebook, but not overall Facebook usage, was significantly correlated with weight dissatisfaction, thin ideal internalisation, and self-objectification [57]. Our study has avoided many of these issues by utilising Facebook likes, rather than status updates or profile pictures. Although technology is growing, as is social media and its associated websites, individuals can be a different person online both physically and socially, which can impact their cognitive and mental functioning in detrimental ways.

The large increase (though due to multicollinearity) in agreeableness in the machine learning model should be inferred cautiously. Agreeableness is characterised by positive social relationships, friendliness, compassion, and cooperativeness [3]. On the Internet, individuals can be whoever they want to be and may thus reinvent themselves as a highly agreeable individual. Other traits, such as conscientiousness, refer to an organised, reliable, and consistent individual who enjoys planning, seeking achievements, and pursuing long-term goals [3]. These characteristics may not be evident through Facebook ‘likes’, as social media websites often do not focus on goals and organisation, but on networking friends and individuals. Facebook ‘likes’ are a basic, discrete digital behaviour that work well with linear regression models. While using natural language processing may allow for a greater understanding of machine learning in social media, it may suffer from significant error due to the complexity of language in statistical analysis [26]. Due to the scope of this study, other aspects of Facebook behaviour were not analysed. Further research into the prediction of traits through machine learning could focus on other aspects of Facebook, such as ‘status updates’, friendship networks, and past events attended, as these online expressions may explain the variables more accurately.

As this is a relatively new area of research, ethical considerations must be addressed. No clear guidelines for conduct in online human subjects research currently exist and thus protocols related to designing online studies, data storage, and analysis of results are scarce, as well as contradictory [24,58]. Using the Internet and Facebook as a research tool poses new ethical dilemmas concerning consent, confidentiality, and competence. The researcher may not be in the same room as the participant, nor have

met them, and thus the reliability and validity of results could be decreased or diminished. For the American population, the American Psychological Association lists three documents with guidelines governing how to conduct research utilising the internet, with the most recent from 2003 [59,60]. However, this is nearly fifteen years old and with the increase in Web 2.0-type websites (i.e., non-static pages), this may not be relevant or sufficient for the current state of Internet-based research. Recommendations from the Association of Internet Researchers (AoIR) Ethics Working Committee state that although no concrete guidelines have been set for internet-based research in America, policies and documents such as the UN Declaration of Human Rights, the Nuremburg Code, the Declaration of Helsinki, and the Belmont Report, apply to all types of research [61]. The basics of these documents are to respect the dignity, autonomy, and rights of the human population and to avoid any possibility of harm. The Australian Psychological Society (APS) takes these fundamentals into account and addresses the key issues of technology, quality, control, and security when dealing with Internet-based research. In Australia, psychologists must abide by the International Guidelines on Computer-Based and Internet Delivered Testing [62] and also abide by the APS ethical guidelines [63]. Accordingly, the British Psychological Society has released Ethics Guidelines for Internet-mediated Research that mirror the APS ethical guidelines [64]. Only three western countries' ethical guidelines have been mentioned as they make up most of the sample for the current study. However, further investigation should inspect the ethical guidelines for other non-westernised countries, as it may be possible to conduct Internet research on these populations. Confidentiality, consent, potential limitations, and security of data collected are perhaps more important in Internet-based research due to the potential of hackers and insecure storing of private information. The ethics behind Internet-based research has not been clear in America due to the dated governing documents. This poses a limitation, as the data collected by the 'myPersonality project' is American-based and consists of predominantly American participants. Future research could limit the sample to Australian participants as the Australian ethical guidelines are comprehensive, though the ethics is still questionable due to the American-based project overall.

In terms of implications for SWB and the FFM of personality, this study creates new avenues of measurement for these constructs, as well as an additional understanding of what they constitute in the online world. Future research could alter the methodology to include regression trees, neural networks, or other algorithms in order to further consider the utility of other machine learning algorithms in the computational social sciences. Greater understanding of the variety of techniques available to psychology researchers, as informed by our data science and computer science colleagues, can only enhance research within the field. With the collaboration of researchers from these fields, a greater knowledge could be built upon to evaluate digital behaviours that are related to certain traits, and individuality can be further investigated using digital data and machine learning techniques.

## 5. Conclusions

The current study found that, by using a machine learning model of Facebook 'likes', high SWB was predicted by high extraversion, openness to experience, conscientiousness, and agreeableness, and low neuroticism. The results further enhance the understanding of machine learning in behavioural sciences and how psychological constructs can be predicted through non self-report methods of measurement. However, the issue of multicollinearity remains when attempting to predict relationships between psychological variables, given that the same digital behaviours are utilised for both independent and dependent variables. As technology use in psychological research continues to develop, it is important for researchers to consider how individuals portray themselves on social media influences how a machine learning algorithm may predict their SWB and personality. Through continual investigation into social media opinion expressions online and their relationship with individual constructs, researchers may be able to develop methods of targeting those at risk of low SWB. In doing so, other associated problems (e.g., depression, body image issues) may be able to be recognised and early intervention provided. Social media websites already use 'cookies' in the browser history to predict the most successful advertisements and promotions, so by using machine

learning-derived predictions for psychological constructs that affect SWB, society could benefit in a way that improves individual health and mentality.

**Author Contributions:** S.C. and J.K. formulated the concept, methodology, and quantitative design of this study, and completed the data curation. A.M. and J.K. performed the data analysis. A.M. conceived the research aims and hypotheses, completed the main investigation, and wrote the initial draft. All authors edited and approved the final manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** This research was supported in part by the Monash eResearch Centre and eSolutions-Research Support Services using the MonARCH HPC Cluster. The authors would also like to acknowledge the myPersonality project for collecting the data and making it available for research use.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Demographic count for participants by country ( $n = 21,122$ ).

Country	Frequency	Country	Frequency
Afghanistan	2	Libya	2
Albania	4	Lithuania	19
Angola	1	Luxembourg	2
Antarctica	1	Macedonia	9
Argentina	31	Madagascar	3
Australia	538	Malaysia	54
Austria	10	Maldives	1
Azerbaijan	1	Malta	5
Bahamas	2	Mexico	89
Bahrain	2	Mongolia	1
Bangladesh	11	Morocco	5
Barbados	2	Mozambique	1
Belarus	3	Myanmar	2
Belgium	35	Namibia	3
Belize	1	Nepal	4
Bermuda	1	Netherlands	50
Bolivia	3	Netherlands	1
Bosnia and Herzegovina	11	Antilles	1
Botswana	1	New Zealand	134
Brazil	40	Nicaragua	5
Brunei	4	Nigeria	6
Bulgaria	24	Norway	26
Cambodia	1	Oman	5
Cameroon	1	Pakistan	27
Canada	728	Palestine	1
Cayman Islands	1	Panama	1
Chile	26	Papua New Guinea	1
China	13	Paraguay	2
Colombia	16	Peru	10
Costa Rica	12	Philippines	224
Croatia	22	Poland	33
Cuba	1	Portugal	42
Cyprus	10	Puerto Rico	20
Czech Republic	16	Qatar	3
		Romania	73



Table A1. Cont.

Country	Frequency	Country	Frequency
Democratic Republic of the Congo	1	Russia	16
Denmark	29	Rwanda	1
Djibouti	1	Saint Helena	1
Dominican Republic	6	Saudi Arabia	11
Ecuador	3	Senegal	1
Egypt	9	Serbia	32
El Salvador	2	Singapore	130
Estonia	14	Slovakia	7
Faroe Islands	1	Slovenia	8
Finland	57	Somalia	1
France	57	South Africa	110
Georgia	3	South Korea	3
Germany	60	Spain	30
Ghana	2	Sri Lanka	4
Greece	26	St Lucia	1
Guatemala	6	Suriname	3
Guinea	1	Sweden	46
Guyana	2	Switzerland	15
Honduras	4	Syria	2
Hong Kong	13	Taiwan	6
Hungary	20	Tanzania	2
Iceland	8	Thailand	23
India	146	The Bahamas	1
Indonesia	96	Tonga	1
Iran	6	Trinidad and Tobago	13
Iraq	3	Tunisia	1
Ireland	109	Turkey	8
Isle of Man	4	Turks and Caicos	1
Israel	13	Uganda	1
Italy	44	Ukraine	8
Ivory Coast	1	United Arab Emirates	20
Jamaica	9	United Kingdom	1401
Japan	33	United States	10,962
Jordan	1	Uruguay	7
Kenya	5	US Virgin Islands	1
Korea	15	Uzbekistan	2
Kuwait	5	Venezuela	16
Laos	1	Vietnam	8
Latvia	2	Zambia	1
Lebanon	4	Zimbabwe	2
Liberia	1	No country stated	4953

## Appendix B

Table A2. Covariance Matrix of observed Five Factor Models ( $n = 13,497$ ).

	Ext	Ope	Agr	Con	Neu
Ext	<0.001				
Ope	<−0.001	<0.001			
Agr	<−0.001	<−0.001	<0.001		
Con	<−0.001	<0.001	<−0.001	<0.001	
Neu	<0.001	<0.001	<0.001	<0.001	<0.001

**Table A3.** Covariance Matrix of machine learning-derived Five Factor Models ( $n = 13,497$ ).

	Ext	Ope	Agr	Con	Neu
Ext	<0.001				
Ope	<0.001	<0.001			
Agr	<−0.001	<−0.001	<0.001		
Con	<−0.001	<0.001	<−0.001	<0.001	
Neu	<0.001	<0.001	<0.001	<0.001	<0.001

## Appendix C

**Table A4.** Correlations between variables for the observed scores ( $n = 21,122$ ).

	SWB	Ext	Ope	Neu	Con	Agr
Ext	0.30	–				
Ope	0.05	0.18	–			
Neu	−0.48	−0.38	−0.07	–		
Con	0.29	0.19	0.01	−0.33	–	
Agr	0.25	0.20	0.08	−0.38	0.18	–

**Table A5.** Correlations between variables for the machine learning-derived scores ( $n = 21,122$ ).

	SWB	Ext	Ope	Neu	Con	Agr
Ext	0.38	–				
Ope	−0.11	−0.19	–			
Neu	−0.56	−0.33	0.12	–		
Con	0.53	0.28	−0.32	−0.49	–	
Agr	0.65	0.23	−0.06	−0.28	0.49	–

## References

- Gosling, S.D.; Sandy, C.J.; John, O.P.; Potter, J. Wired but not weird: The promise of the internet in reaching more diverse samples. *Behav. Brain Sci.* **2010**, *33*, 94–95. [[CrossRef](#)] [[PubMed](#)]
- Kosinski, M.; Wang, Y.; Lakkaraju, H.; Leskovec, J. Mining big data to extract patterns and predict real-life outcomes. *Psychol. Methods* **2016**, *21*, 493–506. [[CrossRef](#)] [[PubMed](#)]
- Kosinski, M.; Stillwell, D.; Graepel, T. Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5802–5805. [[CrossRef](#)] [[PubMed](#)]
- Brickman, P.; Campbell, D.T. Hedonic relativism and planning the good society. In *Adaptation-Level Theory: A Symposium*; Apley, M.H., Ed.; Academic Press: New York, NY, USA, 1971; pp. 287–302.
- Easterlin, R.A. Does economic growth improve the human lot? Some empirical evidence. In *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*; David, P.A., Reder, M.W., Eds.; Academic Press Inc.: New York, NY, USA, 1974; Volume 89, pp. 89–125.
- Duncan, O.D. Does money buy satisfaction? *Soc. Indic. Res.* **1975**, *2*, 267–274. [[CrossRef](#)]
- Kahneman, D.; Diener, E.; Schwarz, N. *Well-Being: Foundations of Hedonic Psychology*; Russell Sage Foundation: New York, NY, USA, 1999.
- Steel, P.; Schmidt, J.; Shultz, J. Refining the relationship between personality and subjective well-being. *Psychol. Bull.* **2008**, *134*, 138–161. [[CrossRef](#)] [[PubMed](#)]
- DeNeve, K.; Cooper, H. The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychol. Bull.* **1998**, *124*, 197–229. [[CrossRef](#)] [[PubMed](#)]
- Hayes, N.; Joseph, S. Big 5 correlates of three measures of subjective well-being. *Pers. Individ. Differ.* **2003**, *34*, 723–727. [[CrossRef](#)]
- Diener, E.; Suh, E.; Oishi, S. Recent findings on subjective well-being. *Indian J. Clin. Psychol.* **1997**, *24*, 25–41.

12. McAdams, D.P.; Pals, J.L. A new big five: Fundamental principles for an integrative science of personality. *Am. Psychol.* **2006**, *61*, 204–217. [[CrossRef](#)] [[PubMed](#)]
13. Kluckhohn, C.E.; Murray, H.A.; Schneider, D.M. *Personality in Nature, Society, and Culture*, 2nd ed.; Knopf: Oxford, UK, 1953.
14. Costa, P.T.; McCrae, R.R. Four ways five factors are basic. *Pers. Individ. Differ.* **1992**, *13*, 653–665. [[CrossRef](#)]
15. Buss, D.M. Social adaptation and five major factors of personality. In *The Five-Factor Model of Personality: Theoretical Perspectives*; Wiggins, J.S., Ed.; Guilford Press: New York, NY, USA, 1996; pp. 180–207.
16. Costa, P.T.; Terracciano, A.; McCrae, R.R. Gender differences in personality traits across cultures: Robust and surprising findings. *J. Pers. Soc. Psychol.* **2001**, *81*, 322–331. [[CrossRef](#)] [[PubMed](#)]
17. Eysenck, H.J.; Eysenck, M. *Personality and Individual Differences: A Natural Science Approach*; Springer: New York, NY, USA, 1985.
18. Buchanan, E.; Aycock, J.; Dexter, S.; Dittrich, D.; Hvizdak, E. Computer science security research and human subjects: Emerging considerations for research ethics boards. *J. Empir. Res. Hum. Res. Ethics* **2011**, *6*, 71–83. [[CrossRef](#)] [[PubMed](#)]
19. Mitchell, T.M. *The Discipline of Machine Learning*; Carnegie Mellon University, School of Computer Science, Machine Learning Department: Pittsburgh, PA, USA, 2006; Volume 9.
20. Feldman, G.; Lian, H.; Kosinski, M.; Stillwell, D. Frankly, we do give a damn: The relationship between profanity and honesty. *Soc. Psychol. Pers. Sci.* **2017**, *8*, 816–826. [[CrossRef](#)] [[PubMed](#)]
21. Baayen, R.H. Data mining at the intersection of psychology and linguistics. In *Twenty-First Century Psycholinguistics: Four Cornerstones*; Cutler, A., Ed.; Taylor & Francis Inc.: New York, NY, USA, 2005; pp. 69–83.
22. Kotsiantis, S.B.; Zaharakis, I.; Pintelas, P. Supervised machine learning: A review of classification techniques. In *Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*; Maglogiannis, I.G., Ed.; IOS Press: Amsterdam, The Netherlands, 2007; Volume 160.
23. Stillwell, D.J.; Kosinski, M. Mypersonality project: Example of successful utilization of online social networks for large-scale social research. *Am. Psychol.* **2004**, *59*, 93–104.
24. 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. *Am. Psychol.* **2015**, *70*, 543–556. [[CrossRef](#)] [[PubMed](#)]
25. Youyou, W.; Schwartz, H.A.; Stillwell, D.; Kosinski, M. Birds of a feather do flock together: Behavior-based personality-assessment method reveals personality similarity among couples and friends. *Psychol. Sci.* **2017**. [[CrossRef](#)] [[PubMed](#)]
26. Youyou, W.; Kosinski, M.; Stillwell, D. Computer-based personality judgments are more accurate than those made by humans. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 1036–1040. [[CrossRef](#)] [[PubMed](#)]
27. Kosinski, M.; Stillwell, D.; Kohli, P.; Bachrach, Y.; Graepel, T. *Personality and Website Choice*; ACM Conference on Web Sciences: New York, NY, USA, 2012.
28. Quercia, D.; Kosinski, M.; Stillwell, D.; Crowcroft, J. Our Twitter Profiles, Our Selves: Predicting Personality with Twitter. In Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), Boston, MA, USA, 9–11 October 2011; pp. 180–185.
29. Reece, A.G.; Danforth, C.M. Instagram photos reveal predictive markers of depression. *EPJ Data Sci.* **2017**, *6*, 15. [[CrossRef](#)]
30. Quercia, D.; Bodaghi, M.; Crowcroft, J. Loosing friends on facebook. In Proceedings of the 4th Annual ACM Web Science Conference, Evanston, IL, USA, 22–24 June 2012; pp. 251–254.
31. Kivran-Swaine, F.; Govindan, P.; Naaman, M. The impact of network structure on breaking ties in online social networks: Unfollowing on twitter. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011; pp. 1101–1104.
32. Lambiotte, R.; Kosinski, M. Tracking the digital footprints of personality. *Proc. IEEE* **2014**, *102*, 1934–1939. [[CrossRef](#)]
33. Grant, S.; Langan-Fox, J.; Anglim, J. The big five traits as predictors of subjective and psychological well-being. *Psychol. Rep.* **2009**, *105*, 205–231. [[CrossRef](#)] [[PubMed](#)]

34. Anglim, J.; Grant, S. Predicting psychological and subjective well-being from personality: Incremental prediction from 30 facets over the big 5. *J. Happiness Stud.* **2016**, *17*, 59–80. [CrossRef]
35. Fujita, F. An Investigation of the Relationship between Extraversion, Neuroticism, Positive Affect and Negative Affect. Master's Thesis, University of Illinois, Champaign, IL, USA, 1991.
36. Lucas, R.E.; Fujita, F. Factors influencing the relation between extraversion and pleasant affect. *J. Pers. Soc. Psychol.* **2000**, *79*, 1039–1056. [CrossRef] [PubMed]
37. Argyle, M.; Lu, L. The happiness of extraverts. *Pers. Individ. Differ.* **1990**, *11*, 1011–1017. [CrossRef]
38. Pishva, N.; Ghalehban, M.; Moradi, A.; Hoseini, L. Personality and happiness. *Procedia-Soc. Behav. Sci.* **2011**, *30*, 429–432. [CrossRef]
39. Gray, J.A. *The Neuropsychology of Emotion and Personality*; Oxford University Press: New York, NY, USA, 1987.
40. Headey, B.; Wearing, A.J. *Understanding Happiness: A Theory of Subjective Well-Being*; Longman Cheshire: Melbourne, Australia, 1992.
41. Kosinski, M.; Bachrach, Y.; Kohli, P.; Stillwell, D.; Graepel, T. Manifestations of user personality in website choice and behaviour on online social networks. *Mach. Learn.* **2014**, *95*, 357–380. [CrossRef]
42. Diener, E.; Emmons, R.A.; Larsen, R.J.; Griffin, S. The satisfaction with life scale. *J. Pers. Assess.* **1985**, *49*, 71–75. [CrossRef] [PubMed]
43. Pavot, W.; Diener, E. Review of the satisfaction with life scale. *Psychol. Assess.* **1993**, *5*, 164–172. [CrossRef]
44. Vassar, M. A note on the score reliability for the satisfaction with life scale: An rg study. *Soc. Indic. Res.* **2008**, *86*, 47–57. [CrossRef]
45. Costa, P.T.; McCrae, R.R. Normal personality assessment in clinical practice: The neo personality inventory. *Psychol. Assess.* **1992**, *4*, 5–13. [CrossRef]
46. Costa, P.T.; McCrae, R.R. The revised neo personality inventory (neo-pi-r). In *The SAGE Handbook of Personality Theory and Assessment*; Boyle, G.J., Matthews, G., Saklofske, D.H., Eds.; The Cromwell Press Ltd.: Trowbridge, UK, 2008; Volume 2, pp. 179–198.
47. McCrae, R.R.; Costa, P.T. A contemplated revision of the neo five-factor inventory. *Pers. Individ. Differ.* **2004**, *36*, 587–596. [CrossRef]
48. Young, M.S.; Schinka, J.A. Research validity scales for the neo-pi-r: Additional evidence for reliability and validity. *J. Pers. Assess.* **2001**, *76*, 412–420. [CrossRef] [PubMed]
49. Boyle, G.J.; Matthews, G.; Saklofske, D.H. *The Sage Handbook of Personality Theory and Assessment: Personality Measurement and Testing*; SAGE Publications: Thousand Oaks, CA, USA, 2008.
50. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017.
51. R Studio Team. Rstudio: Integrated Development for R. Available online: <http://www.rstudio.com/> (accessed on 25 August 2018).
52. Gutierrez, D.D. *Machine Learning and Data Science: An Introduction to Statistical Learning Methods with R*; Technics Publications: Basking Ridge, NJ, USA, 2015.
53. Gignac, G.E.; Szodorai, E.T. Effect size guidelines for individual differences researchers. *Pers. Individ. Differ.* **2016**, *102*, 74–78. [CrossRef]
54. Hareven, T.K. The family as process: The historical study of the family cycle. *J. Soc. Hist.* **1974**, *7*, 322–329. [CrossRef]
55. Easterlin, R.A. Life cycle happiness and its sources: Intersections of psychology, economics, and demography. *J. Econ. Psychol.* **2006**, *27*, 463–482. [CrossRef]
56. Holland, G.; Tiggemann, M. A Systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. *Body Image* **2016**, *17*, 100–110. [CrossRef] [PubMed]
57. Meier, E.P.; Gray, J. Facebook photo activity associated with body image disturbance in adolescent girls. *Cyberpsychol. Behav. Soc. Netw.* **2014**, *17*, 199–206. [CrossRef] [PubMed]
58. Solberg, L.B. Data mining on facebook: A free space for researchers or an irb nightmare. *Univ. Ill. J. Law Technol. Policy* **2010**, *2010*, 311.
59. Kraut, R.; Olson, J.; Manaji, M.; Bruckman, A.; Cohen, J.; Couper, M. Psychological research online: Opportunities and challenges. Report prepared for the american psychology association's taskforce on the internet and psychological research. *Am. Psychol.* **2003**, *59*, 105–117. [CrossRef] [PubMed]
60. Hewson, C. Conducting research on the internet. *PSYCHOLOGIST-LEICESTER-* **2003**, *16*, 290–293.

61. Markham, A.; Buchanan, E. *Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0)*; Association of Internet Researchers: Chicago, IL, USA, 2012.
62. International Test Commission. International guidelines on computer-based and internet-delivered testing. *Int. J. Test.* **2006**, *6*, 143–171. [[CrossRef](#)]
63. Australian Psychological Society. *Ethical Guidelines for Providing Psychological Services and Products Using the Internet and Telecommunications Technologies*; Australian Psychological Society: Melbourne, VIC, Australia, 2014.
64. British Psychological Society. *Ethics Guidelines for Internet-Mediated Research*; British Psychological Society: Leicester, UK, 2017.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).