

A Mental Examination---Using Personality to predict Happiness, Altruism and Health

Minyan Chen

Oregon Episcopal School, Portland, Oregon
6300 SW Nicol Rd, Portland, OR 97223
2nd line of address 15230 SW Sparrow Loop, Beaverton, OR +1(971)401-5645

Abstract. According to professor Jokela, psychologists can know the social functioning of a person only by assessing their Personality traits. However, empirical studies have been focused on building linear regressions between only one facet of personality and Life Satisfaction, Altruism and Health accordingly; also, the accuracy of the prediction remained debatable. In practice, scales online help researchers to get data measurements of participants' information needed in the study. Gradient descent works by building the optimized multiple linear regression to model the relationship of a lot of inputs and a single output; python programs enable researchers to test the accuracy of the predicted output of the regressions. The data was from a preparing study by another group of graduated students from Cambridge University, and it contained information of 1769 participants. By splitting the sample into testing sample (33%) and training sample (67%), three multiple linear regressions were built to model the relationship between 120 Personality items and an average Life Satisfaction score, Altruism score and Health score using the training sample; then, the accuracy of the models was tested using the testing sample. According to the small p-values of correlation between the y-reported and y-predicted for all the three predictions, the probability of getting extreme values was very small, which ensured the reliability of these prediction. According to Cohen's conventions about effect size of correlation in Psychology and another authorized peer research, the Pearson-correlation value of Personality & Life Satisfaction regression shows a very high accuracy of using Personality to predict Life Satisfaction; also, the correlation values for Personality & Altruism and Personality & Health are also above moderate, which indicate nice and acceptable predictability for two regressions.

1 INTRODUCTION

1.1 Background and Literature Review

In the field of positive psychology, numerous theoretical traditions postulate that people's Personality traits play a big role in their life. In the early 1990s, scientist Schmidt and Hunter had proposed in their study that "Personality inventories have rather good validities in the prediction of performance"(Schmidt and Hunter, 1993). Wellbeing and social functioning are very important factors for scientists, doctors, employers and partners to see people's the ability to fulfill their role in work, social activities, and relationships. Thus, scientists all over the world have studied the relationship between personalities and people's well-being and social functioning (e.g., happiness, Health, and Altruism; cite). A broad range of Psychology research has proposed to divide Personality into five traits (the Big Five): extraversion, agreeableness, openness, conscientiousness, and neuroticism. Openness reflects "the ability and tendency to process abstract and perceptual information flexibly and and effectively"

(DeYoung, 2010); Conscientiousness shows "the ability and tendency of individuals to inhibit or constrain their impulses in order to follow rules are pursue non-immediate goals" (Thoms, 1996). Agreeableness links to "people's ability to understand others emotions and intentions as well as empathy" (Roberts, 2007). Extraversion is "the tendency to experience positive emotions, which typically stem from experiences of reward or the promise of reward" (DeYoung, 2010).

Existing research has digged into the relationship between different Personality traits and Life Satisfaction (Happiness), Altruism, and Health. For example, a study by scientist John Brebner and his colleagues has shown that extraversion and neuroticism of Personality traits were very important predictors for predicting Life Satisfaction using multiple regressions (Brebner et al. 1995). Another study in 2000 done by professor Robert F. Krueger and his colleagues postulates that Altruism was positively related to Personality traits such as Social Closeness, Social Potency, and Absorption, and was negatively related to Aggression (F. Krueger et al., 2001). Additionally, in 1995, a study by professor M F. Scheier and M W. Bridges in Carnegie Mellon University suggests that Personality traits such as Hostility,

* Corresponding author: chenj@go.oes.edu

Pessimism, Depression, and Suppression of emotion are related to CHD, and within these traits, Fighting spirit, Pessimism, and Suppression are linked to cancer and AIDS-related diseases (F. Scheier and W. Bridges, 1995). In conclusion, all of the three existing research findings show that there is a strong link between Personality traits and a person's wellbeing and social functioning.

Although existing studies have established a sound relationship between Personality traits and people's well-being and social functioning (e.g., happiness, Health, and Altruism), several problems remain unsolved. First, most of the past studies have been mainly focused on studying the relationship between one of the five Personality traits and happiness, Health and Altruism. This approach ignores the interplay of different Personality traits and its effect on people's life. However, this study proposed that examining Personality as a whole take account of the interactions effects and nuances and can provide more comprehensive insights for understanding the relationship between Personality and people's well-being and social functioning. More importantly, all existing works only focus on using linear regressions to understand the relationship between a Personality trait and people's well-being (e.g., extraversion is positively correlated with happiness), and very few of them go beyond this to explore whether and how scientists can use personalities to forecast people's well-being and social functioning accurately. Building models on personalities to forecast people's well-being and social functioning has great impact on society. For instance, the prediction of mental, physical Health and social functioning based on personalities help doctors to diagnose diseases, help companies to hire the best-fitted employees, and help commercials to target the right customers.

The current study aims to build models to predict well-being (happiness and Health) and social functioning (Altruism) of people based on personalities. Specifically, the study plans to build multiple linear regressions based on a full Personality scale (contains 120 items) that takes into account of all Personality traits, and trains the model to examine the accuracy of the prediction, and provide insights for future social forecasting models.

1.2 Regression Machine Learning Modelling

Regression modelling is a typical and powerful machine learning method to build an association between an outcome and a set of features. The regression model type used in this study is multiple linear regression, which is linear regression with multiple variables. In the multiple linear regression, a dependent target variable Y is modeled as a linear function of several independent variables X_1, X_2, \dots, X_n . The independent variables, also known as inputs of the regression, are factors those help to predict the output. In this study, the inputs are the scores of the 120 questions on the IPIP-NEO test; there are 120 independent variables: X_1, X_2, \dots, X_{120} . The Y values of the three regressions are the average scores on the Life Satisfaction scale, the Happiness score, or the

Altruism scale. Because each regression has 120 independent variables, the graph is not a scatter plot.

The final goal is using Personality information to predict people's Life Satisfaction, Altruism, and Health. In order to achieve the goal, three regressions are built to model the relationship between Personality traits and Life Satisfaction, Altruism and Health; the accuracy of the three models was examined. To begin with, a 2000 participants' data was collected online. The NEO IPIP survey was used to measure people's personalities, the Subjective Wellbeing survey was used to measure people's Life Satisfaction (Happiness), and Altruism values was measured by the Altruistic Personality Scale, and the level Health was self reported using a scale from 1-4. The data was then splitted into two samples: the training sample (66.6% of the data) and the testing sample (with 33.3% of the data). Machine learning was applied to build and test the regression models. Python packages such as numpy, pandas, and matplotlib were installed into the computer to build multiple linear regression models on the training sample to model the relationship between personalities and the three attributes (happiness, Altruism and Health). Descent gradient was used to optimize the regression. After the optimization, the testing sample was put into each of the regression, and the p-values and Pearson correlation values between the predicted outcomes and reported outcomes of the three regressions were calculated by the computer to indicate the accuracy of the three regression models.

2 METHOD

The data used in this study was collected by Li Sai and her colleagues from Cambridge University for another preparing study, and Minyan collaborated with the collectors(Li et al., 2017).

2.1 Participants

2000 participants (1769 of whom completed all of the four surveys; $M_{Age} = 34.5$, $SD = 11.8$; 54% male) from the U.S. through Amazon Mechanical Turk (AMT) joined the study. Among the participants, 76.6% were White/Caucasian, 11.6% were Black/African American, 5.5% were Hispanic/Latino, 2.6% were mixed, 0.2% were Hawaiian or Pacific Islander, 1.1% were American Indian, 1.1% were from Asian Subcontinent, 1.3% were South-East Asian. All participants completed the study online. In accordance with standard AMT wages, each participant was given a US \$1.00 fee for participating. Participants were told that they would conduct surveys for researchers to understand psychological traits and interpersonal relationships.

2.2 Materials and Procedure.

Participants in previous research were instructed to firstly conduct the IPIP-NEO 120 survey online, each item rated on a five-point scale (1 is strongly disagree; 5 is strongly agree). The scores of the 120 items about Personality information are inputs in the multiple linear

regression (X1, X2,..., X120). The survey was from the official website of International Personality Item Pool (IPIP), which was originally created by Wim K. B. Hofstee and his colleagues and students at University of Groningen in the Netherlands (Goldberg et al., 2017). The items and scales were in the public domain. The responses were recorded automatically via the survey program. The internal consistency of the IPIP estimates were 0.85 for extraversion, 0.79 for agreeableness, 0.84 for conscientiousness, 0.88 for neuroticism, and 0.85 for openness to experience.

Then, Life Satisfaction of the participants was measured with the Satisfaction With Life Scale (SWLS). In the system, participants rated five statements about their Life Satisfaction on a seven-point scale (1 is strongly disagree; 7 is strongly agree). The average score of each participant on the five statements was the output of the regression model between Personality & Life Satisfaction (Happiness). The internal consistency of the SWLS estimates was 0.74 (Ortega et al., 2016).

Altruism information of the participants was measured with the Altruistic Personality Scale. In the scale, participants were asked to fill a twenty-item scale designed to measure altruistic tendency by gauging the frequency one engages in altruistic acts primarily toward strangers. They answered on a 5-point scale ranging from Never (0) to Very Often (4) (Rushton, J. P., Chrisjohn, R.D., & Fekken, G. C., 1981). The average score of the twenty statements was the output of the regression model between Personality & Altruism.

To measure participants' level of Health, the participants were asked to self-reported their Health by scaling their state of Health these days (1 = poor, 2 = fair, 3 = good, 4 = very good). The reported score of the participants was the output of the regression model between Personality & Health.

2.3 Statistical Analysis.

In the current study, machine learning models were built to predict a person's Life Satisfaction score average, Altruism score average, and Health score (because there was only one scale for Health) accordingly using their 120 scores from the Personality test.

In general, a machine learning problem considers a set of n samples of data and then tries to predict properties of unknown data. The analysis started by instructing the computer program to randomly split the data into a training set (66.6% of the data) and a testing set (33.3% of the data). Three separate multiple linear regression models were built on the training set, where the 120 items of NEO-IPIP served as predictors, and the averages of the SWL score, the Altruistic Personality Scale score, and the self-reported Health scale score were the outcomes accordingly. Then the accuracy of the three regression models was tested on the testing set. It was important not to test the prediction of an estimator on the data used to fit the estimator, as this would not be evaluating the performance of the estimator on new data. Below are the specific steps of the study.

For predicting each of the averages of the total scores on Life Satisfaction scale, Altruism Scale and Health Scale, the python packages: numpy, pandas, matplotlib, pylab were imported into the computer. Then, the scores of the 120 items in NEO-IPIP were loaded into the system as independent variables. Feature scaling was applied to standardize the range of the independent variables, so that the gradient descent would converge faster. Then, a multiple regression was built to model the relationship between the 120 independent variables on a Personality scale and the average score of the Altruism scale. Gradient descent was applied to find the minimum of the cost (i.e., to minimize the difference between the predicted and the observed outcome, thus finding the optimized coefficients for the 120 variables).

After the multiple linear regression of Personality and Life Satisfaction was built, the accuracy of the model was analyzed on the testing set. The predictors of the testing sample were put into the model to get a predicted outcome (y -predicted). Then, each y -predicted was compared with the reported y value in the testing sample by calculating the Pearson correlation and the p -value. Pearson correlation has a value between +1 and -1, where 1 is total positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation. A small p -value indicates that it is unlikely to observe such a substantial association between the predictor and the response due to chance, which infers that there is a real association between the predictor and the response.

To examine the relationship between Personality and Altruism and the relationship between Personality and Health, the steps above are first using training sample to build the linear regression to model the relationship, and then using the training sample to predict outcomes, and observing the p -value and Pearson correlation were repeated.

3 RESULT

The project originally aimed to build machine-learning models to predict people's well-being (happiness and Health) and social functioning (Altruism) using a full-set of Personality traits. In the project, three multiple linear regressions models were used to model the relationships between Personality and Life Satisfaction, Altruism, and Health, and then examined the accuracy of the three regressions to see whether the prediction is solid .

3.1 An Overview of the Descriptive Statistics

The computer showed that, from the data collection, the average score of the level of satisfaction reported by the participants was 3.006, and the standard deviation of Life Satisfaction was 0.730; the average score of the Altruism reported by the participants was 3.112, and the standard deviation of Altruism was 0.940; the average score of Health reported by the participants was 2.949.

3.2 Accuracy of the Machine Learning Prediction

To examine the accuracy of three models, the inputs of the testing sample (33.3%) were used to get predicted y values, and then the Pearson correlation between the predicted outcome and the self-reported outcome was calculated.

The Pearson-correlation between the predicted SWL and the observed SWL in the testing sample is 0.639 ($p < 0.001$). The R-square of the correlation value is 0.409. The Pearson-correlation between the predicted SWL and the observed SWL in the testing sample is 0.482 ($p < 0.001$), and the R-square of the correlation value 0.234. The Pearson-correlation between the predicted Health and the observed Health in the testing sample is 0.451 ($p < 0.001$), and the R-square of the correlation is 0.204. Via the rule of thumb in Pearson correlation coefficient, all of the three predictions achieved a middle-large effect size.

4 CONCLUSION

The hypothesis of the project was whether Personality can predict a person's Life Satisfaction, Altruism and Health. The current result data provided several insight into the accuracy of using Personality to predict a person's Life Satisfaction, Altruism, and Health. According to the cultural percentage, 76.6% were White/Caucasian, 11.6% were Black/African American, 5.5% were Hispanic/Latino, 2.6% were mixed, 0.2% were Hawaiian or Pacific Islander, 1.1% were American Indian, 1.1% were from Asian Subcontinent, 1.3% were South-East Asian; the participants were from various races. MAge =34.5 suggested that the participants were from different age ranges; 54% male showed that there were approximately equal number of people from both genders. In conclusion, the data was representative to generalize the real world behavior. Also, the training sample and the testing sample were assigned randomly by the computer, so that the conclusion of the prediction is valid. In the experiment, the pearson-correlation of the first regression of Personality and Life Satisfaction was 0.639, which postulated that there was a positive correlation between Personality and Life Satisfaction. The R-square value of correlation indicated that the self-reported Life Satisfaction score changes 40.9% could be explained by the predicted Life Satisfaction score change from the regression. Moreover, the extremely small p -value of the same regression indicated that the probability of getting extreme value was very small, which ensured the reliability of the prediction.

The second regression modeled the relationship between Personality and Altruism. Its pearson correlation was 0.482, which indicated that there was a positive correlation between Personality and Life Satisfaction. The R-square value of the correlation indicated that, the self-reported Altruism score changes 23.4% could be explained by the predicted Altruism score change from the regression. The extremely small p -value of the Personality & Altruism regression ensured the reliability of the prediction.

The last regression modeled the relationship between Personality and Health. The pearson correlation between

Personality and Health was also about 0.451, which indicated that there was a positive correlation between Health and Altruism. The R-square value of the correlation indicated that, the self-reported Health score changes 20.4% could be explained by the predicted Health score change from the regression model. Moreover, the p -value of using Personality to predict Health was also small, which suggested that the prediction results was reliable.

The correlation values were interpreted according to professor Jacob Cohen's conventions in Statistical Power Analysis for the Behavioral Sciences. In psychological research, a correlation coefficient of 0.10 is thought to represent a weak or small association; a correlation coefficient of 0.30 is considered a moderate correlation; and a correlation coefficient of 0.50 or larger is thought to represent a very strong or large correlation (Cohen, 1988). In this study, the correlation value of the first regression (Personality & Life Satisfaction) was 0.639, which represented a very good correlation, thus suggesting that using Personality to predict Life Satisfaction had a very good effect. The correlation values of Personality & Altruism and Personality & Health were also above the moderate correlation coefficient, which suggested that the other two predictions also had more-than-moderate effects.

Additionally, the Pearson-correlation results from another authoritative Psychological paper were used to interpret whether the more-than-moderate effects of the Personality & Altruism and Personality & Health regressions were considered acceptable. According to professor Michael Kosinski, in Psychology, medium-range correlation are considered "good" in Psychology. The Pearson-correlation in Professor Kosinski's paper about Private traits and attributes are predictable from digital records of human behavior were in the medium range, such as 0.3 to 0.5, but they were interpreted as "acceptable effect"(Kosinski, M., Stillwell, D., & Graepel, 2013). The pearson correlation values of Personality & Altruism and Personality & Health were 0.482 and 0.451 (all above the acceptable value in professor Kosinski's research), and they should be nice correlation values to support the accuracy of the Personality predictions.

However, there were still bias in the project. To begin with, according to counselor Darine F. Brown and his colleagues, using questionnaires to be the scales could create distorted results because of sampling error, poorly phrased question, and participants might also reported invalid answers (Brown et al., 1981). Additionally, the Health scale only included one item, and participants might be unaware of their Health status and gave inaccurate self reported value, which might also influenced the model. Lastly, there participation bias might occurred in the experiment as well, because the sample was not big enough, some results might become non-representative because the participants disproportionately possessed certain traits which affected the outcome.

The project successfully uses statistics and Mathematical regression to show that there is a good relationship between Personality and Life Satisfaction,

which supports future studies that aim to build regression between Personality and Life Satisfaction. But because of the limitation of time, the study only retrieved Personality traits from the IPIP Personality test, which included 120 items for a person's Personality traits. The test is quite sophisticated and time-consuming. Future studies can focus on trying to retrieve Personality information in an easier way; instead of asking each of them to fill out a long Personality scale, future studies might focus on using less Personality items to accurately predict attributes including Happiness, Altruism and Health. For example, retrieving enough Personality information from people's online profile or their digital footprints. Moreover, the result shows that there are acceptable relationships between Personality information and Altruism, Personality and Health, and future studies can try other types of multiple regressions to build even more accurate predictions.

The impact of the result was significant, because it successfully showed that psychologists can use Personality information to determine one's social functioning and mental health; people can self diagnose and even predict their mental and physical health using their Personality information at home, and companies can use people's Personality information to recruit best-fitted employees.

REFERENCES

1. Brebner, J., Donaldson, J., Kirby, N., & Ward, L. (1995). Relationships between happiness and Personality. *Personality and Individual Differences*, 19(2), 251-258. doi:10.1016/0191-8869(95)00022-x
2. Brown, D., Hartman, B., & Fuqua, D. (1981). Guidelines for Consumers of Survey Research. *The School Counselor*, 28(4), 279-284. Retrieved from <http://www.jstor.org/stable/23901454>
3. Cohen, J. (1988). *Statistical power analysis for the behavioral science*. New York: Lawrence Erlbaum Associates.
4. DeYoung, C., Hirsh, J., Shane, M., Papademetris, X., Rajeevan, N., & Gray, J. (2010). Testing Predictions From Personality Neuroscience: Brain Structure and the Big Five. *Psychological Science*, 21(6), 820-828. Retrieved from <http://www.jstor.org/stable/41062296>
5. Goldberg, L. et al., (2017). IPIP-NEO. <http://www.personal.psu.edu/cgi-bin/users/j/5/j5j/IPIP/shorttipipneo1.cgi>
6. Krueger, R. F., Hicks, B. M., & McGue, M. (2001). Altruism and Antisocial Behavior: Independent Tendencies, Unique Personality Correlates, Distinct Etiologies. *Psychological Science*, 12(5), 397-402. doi:10.1111/1467-9280.00373
7. Li, S. et al., (in preparation). Efficiency or Equality - A Personality Analysis.
8. Ortega, L. et al., (2016). Satisfaction With Life Scale. <https://hqlo.biomedcentral.com/articles/10.1186/s12955-016-0573-9>
9. Rushton, J. P., Chrisjohn, R.D., & Fekken, G. C. (1981). Altruistic Personality Scale. http://fetzer.org/sites/PROSOCIAL_ORIENTATIO_N_AND_ALTRUISM.pdf
10. Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
11. Scheier, M. F., & Bridges, M. W. (n.d.). Person Variables and Health Personality Predispositions and Acute Psychological States as shared Determinants for Diseases. Retrieved February 13, 2018, from <http://repository.cmu.edu/psychology>
12. Thomas, P., Moore, K., & Scott, K. (1996). The Relationship between Self-Efficacy for Participating in Self-Managed Work Groups and the Big Five Personality Dimensions. *Journal of Organizational Behavior*, 17(4), 349-362. Retrieved from <http://www.jstor.org/stable/2488546>
13. Touzé, P. (2005). Personality AND PREDICTION OF PERFORMANCE IN THE WORKPLACE. *Le Travail Humain*, 68(1), 37-53. Retrieved from <http://www.jstor.org/stable/41507301>
14. Waller, N., & Zavala, J. (1993). Evaluating the Big Five. *Psychological Inquiry*, 4(2), 131-134. Retrieved from <http://www.jstor.org/stable/1448844>
15. Ye, N. (2003). *The Handbook of Data Mining*. Retrieved October 7, 2017, from <https://www.questiaschool.com/library/107581526/the-handbook-of-data-min>