Turkish Journal of Computer and Mathematics Education *Vol. 12 No.14 (2021), 2247-2255*

Research Article

Modelling Relationship between Demographic Data and Personality Traits

Iman Paryudi [a,b], Ahmad Ashari [c], Khabib Mustofa [d]

[a] PhD Student, Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia
[b] Lecturer, Department of Informatics, Universitas Pancasila, Jakarta, Indonesia
[c,d] Lecturer, Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia
[a] paryudi@mail.ugm.ac.id, [b] iman.paryudi@univpancasila.ac.id, [c] ashari@ugm.ac.id, [d] khabib@ugm.ac.id

Abstract: Currently, the most common method to predict personality trait implicitly (Implicit Personality Elicitation) is Personality Elicitation from Text (PET). PET predicts personality traits implicitly based on status written on social media. However, when this method is applied in a recommender system, it has two weaknesses: the obligation to have at least one social media account and the requirement to write status with a certain length. To cope with this shortcoming, we propose a new method to predict personality traits implicitly based on demographic data. A personality model correlating demographic data and personality traits is needed to be able to predict personality traits based on demographic data. In this research, we create 325 models, 65 models each for the five traits (extraversion, agreeableness, conscientiousness, emotional stability, and intellect). To choose the working model, we use the following criteria: the demographic data never change throughout life, a small number of categories at each demographic data, and the model has fairly good accuracy. Based on the criteria, we pick the model that is formed by a combination of age group and gender. There are six age group-gender cohorts in the model: adolescence male, adolescence female, adulthood male, adulthood female, middle-age male and middle-age female. The personality model we obtain is low extraversion for all age group-gender cohorts, high agreeableness and conscientiousness for all age group-gender cohorts. We also find that adolescence male and female have low emotional stability. Meanwhile, the other age group-gender cohorts have high emotional stability. Our model also shows high intellect for adolescence male and middle-age male and low intellect for the other age group-gender cohorts.

Keywords: Implicit personality elicitation, recommender system, demographic data, personality model

1. Introduction

The main shortcoming of a recommender system based on rating-based collaborative filtering is that it cannot recommend accurate items to new users, which is known as the cold-start problem. Researchers make use of personality traits to overcome this weakness. The application of personality traits has three advantages, namely: the system can recommend correct items to new users, it is not influenced by data rating, and it is not bound to a certain domain (Paryudi et al., 2019).

There are two ways to predict someone's personality trait, explicitly and implicitly. The explicit method utilizes a questionnaire to predict one's personality trait. A recommender system applying an explicit method requires its users to fill in a personality questionnaire before becoming a member of that system. The requirement to fill in a questionnaire is regarded as a burdensome and wasting-time task (Tkalčič & Chen, 2015). Because of this problem, the implicit method is used. This method predicts personality traits implicitly from text and therefore it is called Personality Elicitation from Text (PET). The text used to predict personality traits are taken from statuses written in social media accounts like Twitter ((Golbeck, Robles, Edmondson, et al., 2011), (Carducci et al., 2018)); Facebook (Golbeck, Robles, & Turner, 2011); TripAdvisor ((Roshchina et al., 2015), (Di Rienzo & Neishabouri, 2016)); FriendFeed (Celli, 2012); and Weblog (Oberlander & Nowson, 2006).

The infirmity of this method when it is applied in a recommender system is that the users are required to have at least one social media account and write a status with a certain length, which cannot be only one or two words. Hence, a user who does not fulfill this qualification cannot use the PET-based recommender system.

Many researchers have found correlations between demographic data and personality trait, among others, are: age and gender (Soto et al., 2011), race/ethnicity/country ((McCrae et al., 2005), (Schmitt et al., 2007)), sports ((Chong & Mustaffa, 2012), (Steca et al., 2018)), zodiac (Borreli, 2015), blood group (Rogers & Glendon, 2003), color ((Navaro et al., 2013), (Miao, 2017)). Based on these findings, we propose a new method to predict personality traits implicitly based on demographic data to overcome the weakness of the PET method. In this case, we will use several demographic data such as age, gender, tribe/ethnicity, job, hobby, sports, blood group, zodiac, color, and marital status.

To find out the relationship between demographic data and personality traits, we carry out modeling using the Decision Tree method. This paper presents the modeling result.

2. Method

2.1. Participants and Procedure

The research involves 1014 participants from several cities in Indonesia. Of all participants, 56,21% are married and 43,79% are not married.

The data is collected through a questionnaire that is made in the form of Google Form. The questionnaires are distributed to participants via WhatsApp message directly or indirectly (WhatsApp Group).

The questionnaire has two parts: personality trait and demographic data.

a. Personality trait. Participants are required to rate their personality traits by giving a score to each statement. The scoring system utilized in this questionnaire is five Likert scales: 1-Very Inaccurate, 2-Moderately Accurate, 3-Neither Accurate nor Inaccurate, 4-Moderately Accurate, 5-Very Accurate.
   In this survey, we employ big five-based personality questionnaire (Paryudi et al., 2019) and (McCrae & John, 1992). From a variety of big five-based personality questionnaire, we choose to use IPIP 50 questionnaire (Akhtar & Azwar, 2019). It is important to inform readers that, although it is a big five-based questionnaire, IPIP 50 does not use the terms neuroticism and openness instead emotional stability and intellect, respectively. Emotional stability is the opposite of neuroticism.
b. Demographic data. Participants are asked to answer questions on their data such as the city of residence, marital status, birth year, gender, tribe/ethnicity, jobs, hobby, sports, blood group, zodiac, and color.

2.2. Data Preparation

The data that we collect consists of 61 attributes: (1) 50 attributes containing answers (in the form of score) of 50 questions in the IPIP 50 questionnaire. The 50 questions are from 10 questions each from the five traits, (2) City of residence, (3) Marital status, (4) Birth year, (5) Gender, (6) Tribe/ethnicity, (7) Jobs, (8) Hobby, (9) Sports, (10) Blood group, (11) Zodiac, (12) Color.

In this stage, we process the raw data as a preparation for further data processing. Activities performed include:

i. Calculate the total personality score for each trait. The scores collected in the data collection stage are the score for each question. Therefore, the total score for each trait must be calculated.
ii. Convert birth year into age.

According to (Soto et al., 2011), two kinds of analysis must be carried out to check the collected data quality:

(1) Alpha reliabilities. This measure is used to find the internal consistency of the collected data. This measure can be obtained by calculating Cronbach's alpha. From the calculation, the Cronbach’s alpha for each trait are: extraversion 0,801, agreeableness 0,773, conscientiousness 0,844, emotional stability 0,908, and intellect 0,749 with mean 0,815. Based on (StatisticsHowTo, 2021), these alpha values indicate that the internal consistency of the collected data are from acceptable (agreeableness and intellect), good (extraversion and conscientiousness) to excellent (emotional stability). The obtained alpha values are similar to the ones obtained by (Akhtar & Azwar, 2019) they are: extraversion 0,839, agreeableness 0,762, conscientiousness 0,811, emotional stability 0,862, and intellect 0,768.
(2) Interscale correlation. This measure must be calculated to check the presence of self-enhancers, which are participants who exaggerate their personality. The presence of a large number of self-enhancers is characterized by the magnitude of the correlation that is larger than usual. The average correlation obtained by (Soto et al., 2011) is 0,19. A more detailed explanation of self-enhancer is provided at the Data Cleaning stage.

2.3. Exploratory Data Analysis

As previously mentioned, the number of collected data is 1014 data with each record has 61 attributes. Six attributes are added at the Data Preparation stage, namely: extraversion score, agreeableness score, conscientiousness score, emotional stability score, intellect score, and age. At the feature engineering stage, those personality scores will be converted into personality levels. The attribute personality level will be used as a target variable at the modeling stage.

From the quick checking, we find dubious data. First, we find a participant who gives a score of 3 for all questions. Then, we also find participants who give an extreme score, either 1 or 5, for all questions.

To anticipate missing values, all questions in the questionnaire used in this survey are set to be compulsory. Meanwhile, the potential outlier may come from participants who commit self-enhancing. This is because self-enhancers usually exaggerate their scores hence larger than any other participants.

Participants of ages 16 and 17 years old dominate the data each with 7,10% and 6,61%, respectively. The rest are participants from age of 18 to 50 years old. Based on gender, 61,34% are female and 38,66% are male. Meanwhile, two dominant tribes are Java with 67,85% and Sunda with 12,13%. The other tribes are Batak, Minang, China, Arab, and Others.

Regarding jobs, 48,13% of participants are employee, 31,66% unemployed, 8,09% educator, 6,90% security forces, 5,22% entrepreneur. Participants have variety of hobbies such as outdoor 44,58%, art 31,17%, others 24,25%. Sports pursued by participants are strength and agility sports 36,69%, sports with ball 33,63%, water sports 11,05%, brain sports 6,11%, mountain sports 5,23%, and other 7,30%.

For blood group data, 42,21% of participants are of blood group O, 29,19% blood group B, 20,81% blood group A, and 7,79 are of blood group AB. Regarding zodiac, 49,80% of participants are categorized into odd-numbered zodiac and 50,20% are even-numbered zodiac. Meanwhile based on zodiac component, 23,67% are grouped into Fire, 24,26% Water, 26,13% Air, and 25,94% Earth. Lastly, 32,44% of participants favor warm color and 67,56% favor cold color.

2.4. Data Cleaning

The first step in the data cleaning stage is to check double data. Double data is caused by participants who fill in the questionnaire twice with the same answer. From the checking, we find 25 participants who fill in the questionnaire twice. One of the answers is removed from the data.

In the data preparation stage, we calculate the interscale correlation which is a Pearson correlation coefficient between the five traits. The average interscale correlation value we obtained at that time was 0,38, which is considered too high. According to (Soto et al., 2011), the fairly high interscale correlation is due to the presence of a large number of self-enhancers. These self-enhancers will describe themselves as highly extroverted, agreeable, conscientious, emotionally stable, and open to experience. We perform self-enhancer checking to overcome this problem. In this checking, we search for a high personality score (almost maximum to maximum) in all traits. From this activity, we get 94 participants who are self-enhancers and remove them from the data. After the removal, the average interscale correlation reduces to 0,24 which is close to the value obtained by (Soto et al., 2011), 0,19.

From the self-enhancer checking activity, we also find a participant who gives a score of 3 for all questions. This data is considered dubious and hence gets deleted. By the end of this data cleaning stage, the number of remaining data is 894.

2.5. Feature Engineering

We create several new attributes at this stage. The first one is personality level for all traits. There are two categories for this level: high and low. The categorization is carried out by first finding the highest and the lowest scores in each trait. Then, both scores are summed and divided by 2 to get the mean. The scores lower than the mean are categorized as Low. On the opposite, the scores higher than the mean are categorized as High.

Age will be grouped into a new attribute, age group. The grouping is based on age category classification by the Ministry of Health, Republic Indonesia (Muamala Net, 2019) and shown in Table 1. Since the range of age of the participants is between 16 to 50 years old, then there are only three age categories for attribute age group: adolescence, adulthood, and middle age.

Besides creating new attributes, at this stage, we also remove attributes that are not required at the modeling stage. The remaining attributes for the modeling are age group, gender, tribe/ethnicity, job group, hobby group, sports, blood group, zodiac group, zodiac component, color group, and marital status.

Table 1. Age category classification by Ministry of Health, Republic Indonesia (Muamala Net, 2019).

Range of Age	Category
<= 11 years old	Childhood
12 – 25 years old	Adolescence
26 – 45 years old	Adulthood
46 – 65 years old	Middle Age
> 65 years old	Elderly

3. Results dan Discussion

The modeling is performed by means of the decision tree method. The model evaluation is carried out by using cross-validation with k = 10. Since there are five traits, hence we create five personality models. As the dependent attribute (class) is personality level (extraversion level, agreeableness level, conscientiousness level, emotional stability level, and intellect level). And as the independent attribute is either one demographic data or a combination of two demographic data (Table 2). This process creates 65 models for each trait or in total 325 models.

The modelling results exhibits that the extraversion levels on most demographic data and its combination are low. Only some that show high extraversion level. On the opposite of extraversion, the agreeableness level on most demographic data and its combinations are high. Only a few demographic data and its combination that show low agreeableness level. Similar to the agreeableness model, the conscientiousness level on most demographic data and its combinations are also high. The modelling result using emotional stability as the class shows that the high level is still slightly more dominant than the low level. Meanwhile, for intellect, the modelling result is the opposite to emotional stability. In this trait, low intellect level is more dominant than high level.

3.1. Personality Traits

Of the 65 models, we find interesting result regarding the Indonesian personality traits on some models and is explained below.

Figure 1 shows that a low extraversion level is dominant. Therefore, in general, Indonesian is introvert. This finding is consistent with the research of McCrae et al. (2005 in (Paryudi & Nursari, 2020)). They find that the extraversion score of Indonesian is among the lowest in the world together with Moroccan and Nigerian.

Based on researches by McCrae et al. (2005 in (Paryudi & Nursari, 2020)) and Schmitt et al. (2007 in (Paryudi & Nursari, 2020)), the agreeableness score of Indonesian is above average or tends to be high. Our research finds the same phenomenon. Figure 2 prove our claim where almost all model has high agreeableness level.

Like agreeableness, the conscientiousness score of Indonesian based on the aforementioned two researches is also above average (Paryudi & Nursari, 2020). This finding supports our research result where the conscientiousness level of Indonesian tends to be high (Figure 3).

We find more interesting findings in emotional stability trait. These findings are similar to the findings of previous researches. We obtain that the emotional stability level of adolescence is lower than those of adulthood and middle age (Figure 4). In other words, neuroticism level of adolescence is higher than those of adulthood and middle age. This is similar to the work of (Soto et al., 2011) where their participants show high neuroticism scores from a young age until the age of about 30 years old. Afterward, the score continues to go down until old age.

Researches by (Lynn & Martin, 1997), (Donellan & Lucas, 2008), (Denissen et al., 2008), (Chausson, 2010), (Feingold, 1994), (Costa et al., 2001), (Chapman et al., 2007), (Schmitt et al., 2008), (Weisberg et al., 2011), (Vianello et al., 2013) find that the neuroticism score of a female is higher than that of male. Our research finds the same thing. The emotional stability level of a female is lower than the level of a male (Figure 4). In other words, the neuroticism level of a female is higher than the level of a male.

Another interesting finding is that married person has a high level of emotional stability or a low level of neuroticism. The opposite is true for unmarried people (Figure 4). Unfortunately, there is no reference to this finding.

This research also confirms previous research findings associated with intellect or openness. This trait is related to curious, imaginative, broad-minded, broad interest, and like to find new things (McCrae & John, 1992), (Cantador et al., 2013).

(Soto et al., 2011) find that the openness trend from a young age to old age tends to be flat. We also find a similar trend where there is no intellect level change from adolescence to middle age (Figure 5).

Figure 5 also exhibits that the intellect level of a male is higher than the level of a female. This finding corresponds to the finding that female is higher in extraversion, agreeableness, conscientiousness, and neuroticism than a male. The only trait where a male is higher than a female is openness (Soto et al., 2011). The reason for this is that a male loves to use logic than feeling as is usually done by a female (Feingold, 1994), (Costa et al., 2001).

Openness also encompasses a degree of intelligence, curiosity, and creativity when someone meets new things. This is probably the reason why we find participants, who love brain sports like chess and bridge, has an intellect level that is higher than the other participants (Figure 5).

3.2. Personality Model

The goal of this research is to create a personality model relating demographic data and personality traits. The model is chosen from the models previously created. To choose the working model, we use the following criteria: (1) The demographic data do not change throughout life, (2) A small number of categories in each demographic data. This criterion is meant to get a simple model, (3) Has fairly high accuracy.

Table 2. List of demographic data and a combination of demographic data that are used in the modeling.

No.	Demographic Data	No.	Demographic Data
1	Age Group	34	Tribe/Ethnicity and Blood Group
2	Gender	35	Tribe/Ethnicity and Zodiac Group
3	Tribe/Ethnicity	36	Tribe/Ethnicity and Zodiac Component

4	Job Group
5	Hobby Group
6	Sports
7	Blood Group
8	Zodiac Group
9	Zodiac Component
10	Color Group
11	Marital Status
12	Age Group and Gender
13	Age Group and Tribe/Ethnicity
14	Age Group and Job Group
15	Age Group and Hobby Group
16	Age Group and Sports
17	Age Group and Blood Group
18	Age Group and Zodiac Group
19	Age Group and Zodiac Component
20	Age Group and Color Group
21	Age Group and Marital Status
22	Gender and Tribe/Ethnicity
23	Gender and Job Group
24	Gender and Hobby Group
25	Gender and Sports
26	Gender and Blood Group
27	Gender and Zodiac Group
28	Gender and Zodiac Component
29	Gender and Color Group
30	Gender and Marital Status
31	Tribe/Ethnicity and Job Group
32	Tribe/Ethnicity and Hobby Group
33	Tribe/Ethnicity and Sports
37	Tribe/Ethnicity and Color Group
38	Tribe/Ethnicity and Marital Status
39	Job Group and Hobby Group
40	Job Group and Sports
41	Job Group and Blood Group
42	Job Group and Zodiac Group
43	Job Group and Zodiac Component
44	Job Group and Color Group
45	Job Group and Marital Status
46	Hobby Group and Sports
47	Hobby Group and Blood Group
48	Hobby Group and Zodiac Group
49	Hobby Group and Zodiac Component
50	Hobby Group and Color Group
51	Hobby Group and Marital Status
52	Sports and Blood Group
53	Sports and Zodiac Group
54	Sports and Zodiac Component
55	Sports and Color Group
56	Sports and Marital Status
57	Blood Group and Zodiac Group
58	Blood Group and Zodiac Component
59	Blood Group and Color Group
60	Blood Group and Marital Status
61	Zodiac Group and Color Group
62	Zodiac Group and Marital Status
63	Zodiac Component and Color Group
64	Zodiac Component and Marital Status
65	Color Group and Marital Status

Figure 1. Relationship between demographic data and a combination of demographic data with extraversion level (Vertical line cell: high, white cell: low, black cell: not used).

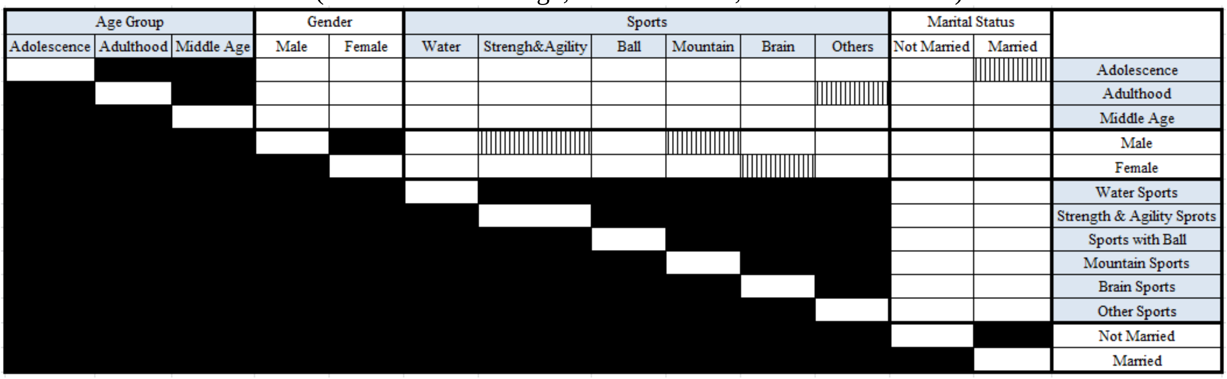


Figure 2. Relationship between demographic data and a combination of demographic data with agreeableness level (Vertical line cell: high, white cell: low, black cell: not used).

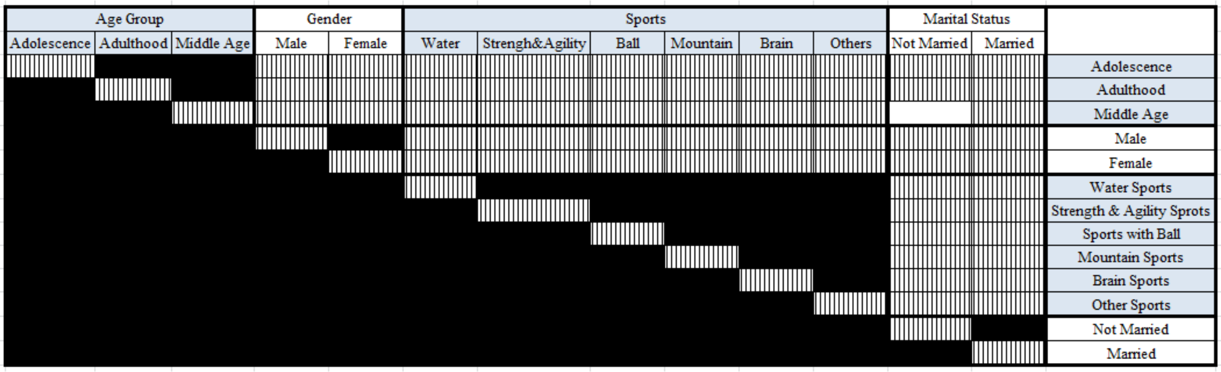


Figure 3. Relationship between demographic data and a combination of demographic data with conscientiousness level (Vertical line cell: high, white cell: low, black cell: not used).

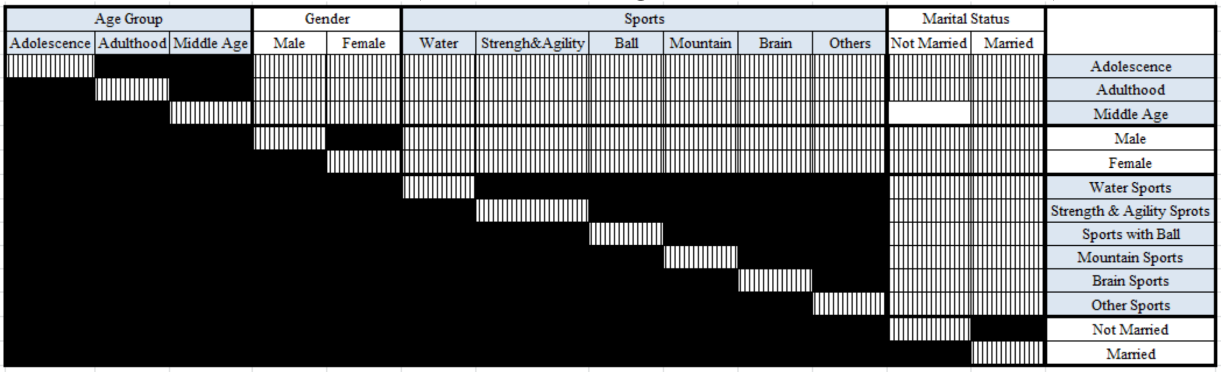


Figure 4. Relationship between demographic data and a combination of demographic data with emotional stability level (Vertical line cell: high, white cell: low, black cell: not used).

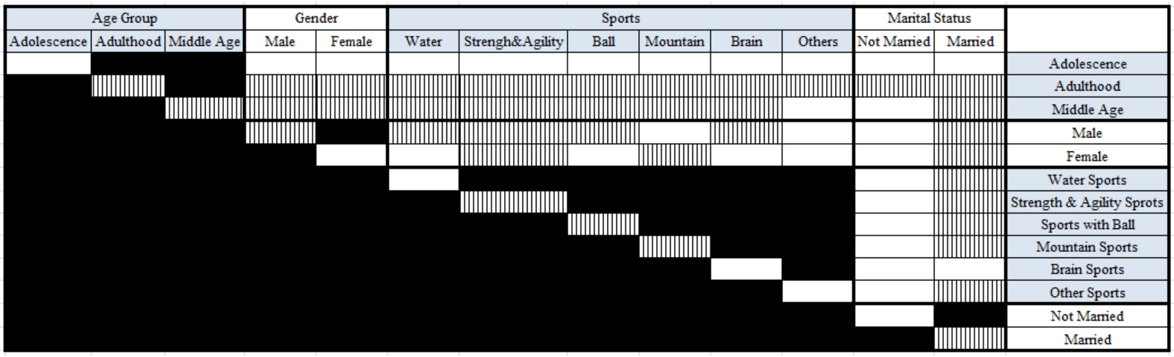
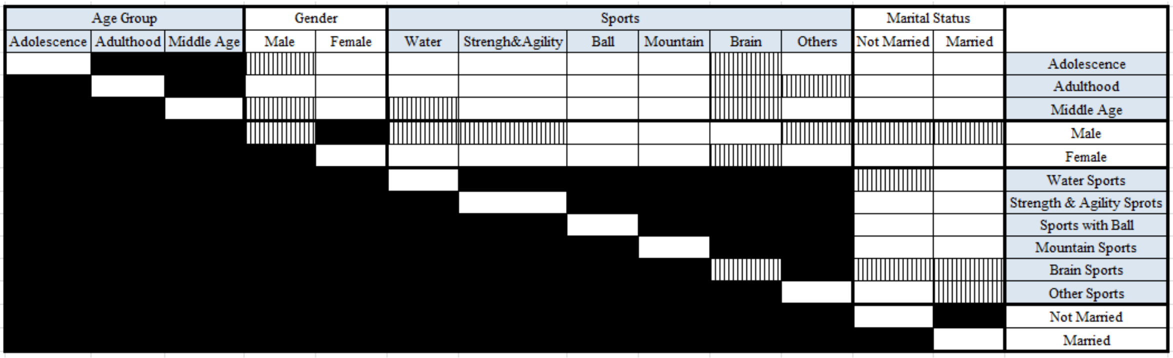


Figure 5. Relationship between demographic data and a combination of demographic data with intellect level (Vertical line cell: high, white cell: low, black cell: not used).



From all models, the model that meets all criteria is the one created based on a combination of age group and gender. The reasons are: (1) the demographic data utilized in the model does not change throughout life, they are: birth year (for age group) and gender. (2) the number of categories for age group is three categories and two categories for gender. This number is small enough to create a simple model. (3) the accuracies of this model for extraversion, agreeableness, conscientiousness, emotional stability, and intellect are 60,5%; 86,8%; 85,0%, 68,1%, and 54,3%, respectively. These values are quite high since they are among the best accuracy of each trait. Several researchers obtain similar model accuracies when predicting personality using PET they are includes Argamon et al (2005 in (Oberlander & Nowson, 2006)) who get average accuracy of about 58%. Model accuracies obtained by (Mairesse et al., 2007) are extraversion 56%, agreeableness 56%, conscientiousness 56%, emotional stability 58%, and openness 63%. (Celli, 2012) get average accuracy of 63,1%. Meanwhile (Di Rienzo & Neishabouri, 2016) obtain the accuracies of each trait are extraversion 100%, agreeableness 100%, conscientiousness 80%, neuroticism 50%, and openness 50%. Another reason to choose a combination of age group and gender is that (Soto et al., 2011) find that there is a very close relationship between age-gender and personality score.

Figure 6 exhibits how the relationship between age group and gender with the personality level of the five traits. From this, we can create a model as shown in Table 2. This is the working model that will be applied in the personality-based recommender system to be built. The model is used to predict user's personalities based on age group and gender.

Figure 6. Correlation between age group and gender with extraversion level, agreeableness level, conscientiousness level, emotional stability level, and intellect level (Red: high, white: low).

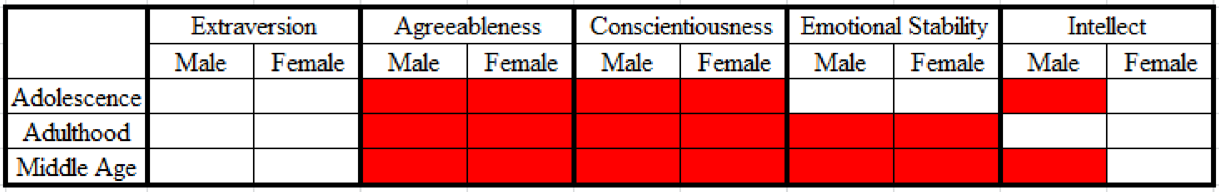


Table 2. Personality model (E: Extraversion, A: Agreeableness, C: Conscientiousness, ES: Emotional Stability, I: Intellect).

	E	A	C	ES	I
Adolescence Male	Low	High	High	Low	High
Adolescence Female	Low	High	High	Low	Low
Adulthood Male	Low	High	High	High	Low
Adulthood Female	Low	High	High	High	Low
Middle Age Male	Low	High	High	High	High
Middle Age Female	Low	High	High	High	Low

4. Conclusion

We create 65 personality models for each trait (extraversion, agreeableness, conscientiousness, emotional stability, and intellect). We set three criteria to choose one model, namely: the demographic data do not change throughout life, a small number of categories in each demographic data, and has fairly high accuracy. The model that meets all criteria is the one with a combination of age group and gender. The personality model we obtain is low extraversion for all age group-gender cohorts, high agreeableness and conscientiousness for all age group-gender cohorts. We also find that adolescence male and female have low emotional stability. Meanwhile, the other age group-gender cohorts have high emotional stability. Our model also shows high intellect for adolescence male and middle-age male and low intellect for the other age group-gender cohorts.

We find interesting results that confirm previous studies. First is that the extraversion level of Indonesian tends to be low. In opposite, the level of agreeableness and conscientiousness of Indonesian tend to be high. The emotional stability level of adolescence is lower (neuroticism level is higher) than that of adulthood and middle age. Besides that emotional stability level of a female is lower (neuroticism level is higher) than the level of a male. For intellect, the intellect level of a male is higher than that of a female. We also find that the intellect level tends to be flat from adolescence to middle age.

References

Akhtar, H., & Azwar, S. (2019). Indonesian Adaptation and Psychometric Properties Evaluation of the Big Five Personality Inventory: IPIP-BFM-50. *Jurnal Psikologi*, *46*(1), 32–44. https://doi.org/10.22146/jpsi.33571

Borreli, L. (2015). *How Birth Month Influences Personality Traits: The Ironic Science Of Astrological Signs*.

Cantador, I., Fernández-Tobías, I., & Bellogín, A. (2013). Relating Personality Types with User Preferences Multiple Entertainment Domains. *EMPIRE 2013*. https://www.researchgate.net/publication/283270671

Carducci, G., Rizzo, G., Monti, D., Palumbo, E., & Morisio, M. (2018). TwitPersonality: Computing personality traits from tweets using word embeddings and supervised learning. *Information (Switzerland)*, *9*(5), 1–20. https://doi.org/10.3390/info9050127

Celli, F. (2012). Unsupervised Personality Recognition for Social Network Sites. *ICDS 2012, The Sixth International Conference on Digital Society*, *c*, 59--62.

Chapman, B. P., Duberstein, P. R., Sörensen, S., & Lyness, J. M. (2007). Gender differences in Five Factor Model personality traits in an elderly cohort. *Personality and Individual Differences*, *43*(6), 1594–1603. https://doi.org/10.1016/j.paid.2007.04.028

Chausson, O. (2010). *GENDER, PERSONALITY AND FILM PREFERENCES Assessing The Impact Of Gender And Personality On Film Preferences*.

Chong, R. N. R., & Mustaffa, M. M. S. (2012). *Relationship Between Sport Preferences and Personality Among Athletes in Malaysian Games 2010*. *March*.

Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, *81*(2), 322–331. https://doi.org/10.1037/0022-3514.81.2.322

Denissen, J. J. A., Geenen, R., Van Aken, M. A. G., Gosling, S. D., & Potter, J. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of Personality Assessment*, *90*(2), 152–157. https://doi.org/10.1080/00223890701845229

Di Rienzo, A., & Neishabouri, A. (2016). Recommendations with personality traits extracted from text reviews. In *Studies in Computational Intelligence* (Vol. 616, pp. 355–364). Springer Verlag. https://doi.org/10.1007/978-3-319-25017-5_33

Donellan, M. B., & Lucas, R. E. (2008). Age Differences in the Big Five Across the Life Span: Evidence from Two National Samples. *Psychology and Aging*, *23*(3), 558–566. https://doi.org/10.1037/a0012897.Age

Feingold, A. (1994). Gender Differences in Personality: A Meta Analysis. *Psychological Bulletin*, *116*(3), 429–456. https://doi.org/10.1176/ajp.152.4.579

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting personality from twitter. *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, 149–156. https://doi.org/10.1109/PASSAT/SocialCom.2011.33

Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. *Conference on Human Factors in Computing Systems - Proceedings*, 253–262. https://doi.org/10.1145/1979742.1979614

Lynn, R., & Martin, T. (1997). Gender differences in extraversion, neuroticism, and psychoticism in 37 nations. *Journal of Social Psychology*, *137*(3), 369–373. https://doi.org/10.1080/00224549709595447

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, *30*, 457–500. https://doi.org/10.1613/jair.2349

McCrae, R. R., & John, O. P. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McCrae, R. R., Terracciano, A., Leibovich, N. B., Schmidt, V., Shakespeare-Finch, J., Neubauer, A., De Fruyt, F., Munyae, M., Flores-Mendoza, C., Dahourou, D., Ayearst, L. E., Trobst, K. K., Alcalay, L., Simonetti, F., Bratko, D., Marušić, I., Høřbíčková, M., Lykke Mortensen, E., Henrik Jensen, H., … Djuric Jocic, D. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology*, *89*(3), 407–425. https://doi.org/10.1037/0022-3514.89.3.407

Miao, R. (2017). *Colour preference*.

Muamala Net. (2019). *Kategori Umur Menurut WHO & Depkes yang Belum Banyak Diketahui Masyarakat*. https://muamala.net/kategori-umur-menurut-who/

Navaro, R. T., Cadorna, E. A., Llanes, L. P., & Esta, M. (2013). Color Preferences and Personality Traits of CAS Faculty : Gender Differences. *UNP Research Journal*, *XXII*(December), 110–127.

Oberlander, J., & Nowson, S. (2006). Whose thumb is it anyway? Classifying author personality from weblog text. *Proceedings of the COLING/ACL*, 627–634.

Paryudi, I., Ashari, A., & Tjoa, A. M. (2019). Personality Estimation using Demographic Data in a Personality-based Recommender System: A Proposal. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3366030.3366098

Paryudi, I., & Nursari, S. R. C. (2020). Graph-based Personality Modeling for Personality-based Recommender System : Do Race and Questionnaire Type Affect Model Testing Result ? *International Journal of Advanced Science and Technology*, *29*(3), 3427–3440.

Rogers, M., & Glendon, A. I. (2003). Blood type and personality. *Personality and Individual Differences*, *34*, 1099–1112.

Roshchina, A., Cardiff, J., & Rosso, P. (2015). TWIN: Personality-based Intelligent Recommender System. *Journal of Intelligent and Fuzzy Systems*, *28*(5), 2059–2071. https://doi.org/10.3233/IFS-141484

Schmitt, D. P., Allik, J., McCrae, R. R., Benet-Martínez, V., Alcalay, L., Ault, L., Austers, I., Bennett, K. L., Bianchi, G., Boholst, F., Borg Cunen, M. A., Braeckman, J., Brainerd, E. G., Caral, L. G. A., Caron, G., Martina Casullo, M., Cunningham, M., Daibo, I., De Backer, C., … Sharan, M. B. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, *38*(2), 173–212. https://doi.org/10.1177/0022022106297299

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why Can't a Man Be More Like a Woman? Sex Differences in Big Five Personality Traits Across 55 Cultures. *Journal of Personality and Social Psychology*, *94*(1), 168–182. https://doi.org/10.1037/a0014651

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age Differences in Personality Traits From 10 to 65: Big Five Domains and Facets in a Large Cross-Sectional Sample. *Journal of Personality and Social Psychology*, *100*(2), 330–348. https://doi.org/10.1037/a0021717

StatisticsHowTo. (2021). *Cronbach's Alpha: Simple Definition, Use and Interpretation*. https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/cronbachs-alpha-spss/

Steca, P., Baretta, D., Greco, A., D'Addario, M., & Monzani, D. (2018). Associations between personality, sports participation and athletic success. A comparison of Big Five in sporting and non-sporting adults. *Personality and Individual Differences*, *121*(October 2017), 176–183. https://doi.org/10.1016/j.paid.2017.09.040

Tkalčič, M., & Chen, L. (2015). Personality and Recommender Systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (Second Edi, pp. 715–739). Springer. https://doi.org/10.1007/978-1-4899-7637-6

Vianello, M., Schnabel, K., Sriram, N., & Nosek, B. A. (2013). Gender Differences in Implicit and Explicit Personality Traits. *Personality and Individual Differences*. https://doi.org/10.2139/ssrn.2249080

Weisberg, Y. J., De Young, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in Psychology*, *2*(AUG). https://doi.org/10.3389/fpsyg.2011.00178