Creating Personality and Preference Models based on Demographic Data for Personality-based Recommender System for Fashion using Decision Tree and Association Rule

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: Several researchers have used posts on social media to estimate the personality traits of the authors. A personality-based recommender system that applies the method to predict the user's personality traits requires the user to write a status of a certain length. The problem with this is that not everyone writes a status on their social media accounts. Therefore, such a recommender system cannot be used for everyone. To solve this problem, we propose a new method of indirectly predicting personality traits which is based on demographic data. To be able to do this, a personality model was needed that relates demographic data to personality traits. As many as 325 personality models were created, of which there were 65 models for each of the following traits namely, agreeableness, emotional stability, intellect, extraversion, conscientiousness. We used three criteria to select the model to be used, namely demographic data that does not change in the course of one's life, it does not have too many categories and the model has quite good accuracy. Based on the above criteria, we chose a model consisting of a combination of age group and gender. Another reason for choosing this model was the findings of previous researchers which state that there is a very close relationship between ages - gender and personality traits. The personality model reveals that each age group – gender cohort has specific personality except for adulthood female and middle age female who have the same personality. To be able to recommend certain items to users of the recommender system, one more model was needed, namely a preference model that connects personality traits with preferences for fashion styles. The preference model shows that only male with low emotional stability and high intellects likes Natural and Masculine fashion style. Meanwhile, male and female with other personalities like Elegant Chic and Natural fashion style.
Keywords: Demographic data, personality-based recommender system, personality trait model, preference model

1. Introduction

The cold-start problem faced by rating-based collaborative filtering forced researchers to use personality traits to solve the problem. The advantage of using personality traits in a recommender system is that the system can provide accurate recommendations as soon as the user becomes a member of the system. Another advantage of personality traits is that they are not affected by rating data. In addition, another advantage is that personality traits are not tied to a particular domain (Paryudi et al., 2019). The latter is important if the recommender system is to use several domains at once such as music, movies, and books.

Before a system can use personality traits to recommend items to users, it must estimate the users' personality traits. The researcher used explicit and implicit methods to estimate user personality traits. Since the explicit method (direct prediction) requires the users to fill out a questionnaire to be able to estimate their personality traits, this method was considered time-consuming and burdensome (Tkalcic & Chen, 2015 in (Paryudi et al., 2019). The implicit method offered a solution to the drawbacks of the explicit method because the implicit method predicts personality traits indirectly. This method predicts personality traits from posts written by a user on social media (personality elicitation from a text). Some social media that have been used to estimate personality traits include Facebook, FriendFeed, TripAdvisor, Twitter, and Weblog ((Golbeck, Robles, & Turner, 2011); (Celli, 2012); (Roshchina et al., 2015); (Di Rienzo & Neishabouri, 2016); (Golbeck, Robles, Edmondson, et al., 2011); (Carducci et al., 2018); (Oberlander & Nowson, 2006)). The weakness of such implicit method when applied to a recommender system is the need to have a social media account with a status on that account. This limits the number of users of this kind of recommender system.

To overcome the above problem, (Paryudi et al., 2021) have offered a new method for predicting personality traits indirectly based on demographic data. This proposal is based on the findings of previous researchers regarding the relationship between personality traits and demographic data. (Soto et al., 2011) found a correlation between gender - age and personality traits. Meanwhile, findings regarding the relationship between race/ethnicity/country have been reported by (Schmitt et al., 2007) and (McCrae et al., 2005). Sports, zodiac signs, blood types, and colors are also known to have a relationship with personality traits based on research results from

(Chong & Mustaffa, 2012), (Steca et al., 2018), (Borreli, 2015), (Rogers & Glendon, 2003), (Navaro et al., 2013), (Miao, 2017).

To be able to apply this new method to a recommender system, two types of models are needed, namely the personality model and the preference model. The personality model relates demographic data to personality traits. The demographic data we used in modeling personality traits are color, zodiac sign, blood type, sport, gender, age, ethnicity, hobbies, occupation, and marital status. Meanwhile, the preference model connects personality traits with a fondness for fashion. This paper presents the modeling results of those two models.

2. Research Method

There were 1014 respondents involved in this study. 56.21% of them were/had once been married and the rest were unmarried. The respondents came from various cities in Indonesia. We collected data using a questionnaire created using Google Forms and distributed via WhatsApp messages. The questionnaire is divided in three sections, namely personality traits, demographic data, and preferences.

We use the IPIP 50 personality trait questionnaire which had been translated into Indonesian (Akhtar & Azwar, 2019). IPIP 50 is a questionnaire based on the big five. As the name suggests, there are five factors/traits in the big five, namely extraversion, agreeableness, conscientiousness, neuroticism, and openness. However, the IPIP 50 questionnaire does not use the terms neuroticism and openness, but instead it uses emotional stability (the opposite of neuroticism) and intellect.

The demographic data that we collected in the survey included city of residence, marital status, year of birth, gender, ethnicity, occupation, hobbies, sports, blood type, zodiac sign, and color.

Meanwhile, in the preference section, respondents were asked to choose an example of their preferred clothing from seven styles of clothing. The seven fashion styles are classic, creative, dramatic, elegant chic, feminine, natural, and rebellious (Maxfield, 2017). The questionnaire provided 15 samples of clothing for each style of clothing so that there were a total of 105 examples of clothing. Respondents were allowed to choose all examples if they liked them all. However, they were also allowed to choose none at all. There were specific notes for feminine fashion style. The term feminine was used in this name of the fashion style as it was originally designed for women. However, when applied to men, it does not mean that the men are dressed in feminine clothing like a woman. For men, of course, it means masculine. The same is true for the term elegant chic. Chic is often associated with women. When used by men, it refers to a masculine elegant chic style.

The data collected had 68 attributes. These attributes were: (a) City of residence, (b) Year of birth, (c) Marital status, (d) Gender, (e) Occupation, (f) Sports, (g) Ethnicity (h) Blood type, (i) Color, (j) Hobbies, (k) Zodiac sign, (l) 50 attributes that contains score to 50 questions on the IPIP 50, (m) 7 attributes containing data on fashion preferences.

The data is processed, firstly, to convert the year of birth into respondent's age. The second is to calculate the total score for each trait. This is because in the data collected during the survey using a questionnaire, a score was given for each question. The third is to determine which clothing samples were chosen by the respondents and how many items were selected in each fashion style. These chosen items are useful when the system recommends items to a user because the recommendations are based on similar preferences with their nearest neighbors (collaborative filtering), whereas the number of selected items is used to determine the level of preference for a fashion style. The level of preference is WEAK when the number of choices is 1 – 5 items; MODERATE when the number of choices is 6 – 10 items. Meanwhile, the preference is STRONG if the number of choices is 11-15 items. In addition, three favorite fashion styles will be determined based on the number of items selected. Three fashion styles with the highest number of selected items will be selected as three favorite fashion styles.

The internal consistency of the data we collected ranged from acceptable to excellent (StatisticsHowTo, 2021) with the Cronbach alpha values of agreeableness, emotional stability, intellect, extraversion, conscientiousness being 0.801, 0.773, 0.844, 0.908, and 0.749, respectively.

38.66% of all respondents involved in the survey were men while 61.34% were women. In terms of age, most of the respondents were aged 16 and 17 years. The percentage of respondents aged 16 years is 7.10%, while that of respondents aged 17 years is 6.61%. Regarding blood type, 42.21% of the respondents had blood type O, 29.19% blood type B, 20.81% blood type A, and 7.79% blood type AB.

Respondents who took part in the survey came from a number of ethnic groups in Indonesia, namely Javanese, Sundanese, Minang, Batak, Arabic, Chinese, and others. A majority were Javanese with a percentage of 67.85% and Sundanese 12.13%. Most of the respondents were employees (48.13%), while others were unemployed (31.66%), educators (8.09%), security officers (6.90%), and entrepreneurs (5.23%). Regarding favorite colors, warm colors were liked by 32.44% of respondents and cool colors were liked by 67.56% of respondents.

Regarding the zodiac signs, 49.80% of respondents were included in the odd numbered zodiac group and the rest were included in the even numbered zodiac group. In addition to number, the zodiac is also grouped based on components (Air, Fire, Earth, and Water). There were 26.13% respondents who have an air component, 23.67% a fire component, 25.94% earth component, and 24.26% water component. We also surveyed favorite sports with the following results: 36.69% liked strength and agility sports, 33.63% liked sports with balls, 11.05% liked water sports, 6.11% liked brain sports, 5, 23% liked mountain sports, and 4.73% liked other sports. The hobbies that were selected by respondents were outdoor activity hobbies 44.58 %, art 31.17%, and other hobbies 24.25%.

To ensure that the data is free of errors, we conducted several checks on the data, namely: double data (respondents who filled out the questionnaire more than once with the same answer), self-enhancer (respondents who exaggerated their answers), and data that did not make sense.

Checking of double data had to be done because the use of a questionnaire created using Google Forms enables a respondent to respond more than once. From this check, we found 25 duplicate data. To clean the data, one of the data was deleted. Checking the self-enhancer could be done by checking the value of the interscale correlation, namely the value of the Pearson correlation coefficient between attributes. At the initial stage, the interscale correlation obtained was 0.38. This value is higher than the value obtained by (Soto et al., 2011) which was 0.19. According to them, this was due to the large number of respondents who scored higher than they should have on all traits. Respondents like this are called self-enhancers. Therefore, self-enhancers will have high levels of personality traits for all traits (high extraversion, high agreeableness, high conscientiousness, high neuroticism, and high openness). 94 self-enhancers were found and then deleted. The interscale correlation value decreased to 0.24 after the self-enhancer data was deleted. The final check was to check the data that did not make sense. The dubious data that we found was respondent who answered all questions with a score of 3. Accordingly, this data was deleted. The amount of data remaining after this checking stage was 894.

To be able to do the modeling, we created a number of new attributes. Among the new attributes are personality levels for five personality traits, namely agreeableness level, emotional stability level, intellect level, extraversion level, and conscientiousness level. There are two categories of levels in personality traits, namely low and high. The ways of classifying levels are as follows: (1) looking for the lowest and highest scores in each trait, (2) finding the average by adding up the two scores and then dividing the total by two, (3) classifying scores that are smaller than the average into low level and classifying higher-than-average scores into high levels.

In addition to adding attributes, we also removed attributes that were not used in the modeling. The remaining attributes that were used in the modeling stage were gender, ethnicity, age group, occupational group, sport, hobby group, color group, zodiac component, zodiac group, blood type, marital status, extraversion level, agreeableness level, conscientiousness level, emotional stability level, intellect level, favorite fashion style 1, favorite level 1, favorite fashion style 2, favorite level 2, favorite fashion style 3, favorite level 3.

3. Results and Discussion

In this study, we created two models, namely the personality model and the preference model. The personality model correlates demographic data with personality traits, while the preference model connects personality traits with a preference for clothing styles.

3.1. Personality Traits Modeling

The personality model was created using the decision tree method. In this model, the dependent and independent attributes are the level of personality traits and demographic data, respectively. We used one demographic data and a combination of two demographic[1] data as independent attributes. 65 models were generated for each trait. For the evaluation of the model, a 10-fold cross-validation method was used.

There were three criteria for selecting the working model from a number of models made. These criteria were: unchanged demographic data in the course of life, not too many categories in the demographic data and the accuracy is quite high.

The first criterion was needed so that the model could apply to everyone forever. It has been found that there is a strong correlation between the need for social roles and changes in personality traits (Heller et al., 2009). For example, when someone is carrying out a social role in which he/she is required to be conscientious, this will result in that person's conscientiousness level becoming high. On the other hand, when in the next phase of life,

[1] From 11 demographic data (gender, ethnicity, age group, occupational group, sports, hobby group, color group, zodiac component, zodiac group, blood type, marital status), we combine two demographic data, for instance gender –ethnicity, gender – sports, etc. This activity yields 54 such combinations. Therefore in total, there are 65 demographic data and their combinations that are used in the modeling.

he/she takes on other social roles that do not require conscientious behavior, his/her conscientiousness level will be low. (Heller et al., 2009) state that one example of a social role is that of an employee or working in a certain workplace. Thus, a change in social role can be caused by a change in the workplace. In this case, personality traits change due to job demands. Due to the demands of work in one workplace, the conscientious level can be high, but in another, it can be low. The results of the study indicate that changes in a person's personality traits can be explained by looking at the changes in the social roles he has gone through. Therefore, social roles are often used as a way to predict personality traits (Ozer & Benet-Martínez, 2006).

The explanation above shows that a problem of using job as a personality trait model occurs when a user of a recommender system has just changed professions from job A to job B which demands different personality traits. When asked about his job, the user answered B. Thus, the recommendation system predicted his personality trait based on job B, even though at that time the influence of his previous job, namely job A on his personality trait, was still very strong. This raises two problems. The first problem is that the prediction of personality traits is incorrect. And because personality traits also affect preference ((Cantador et al., 2013); (Hu, 2010), (Rentfrow & Gosling, 2003); (Rentfrow et al., 2011); (Kraaykamp & van Eijck, 2005); (Chausson, 2010); (Chen et al., 2013)), when given a recommendation, the user does not really like the recommended items. Incorrectly predicting this preference is the second problem.

Based on this, demographic data which can change in the course of a person's life should not be used as a model. Other demographic data included here were hobbies, sports, color, and marital status. So that the demographic data that meet the criteria are only age, gender, ethnicity, blood type, and zodiac sign, either alone or in combination. There is an exception for age which increases with time. In this case, age is considered unchanged because what is used here is the year of birth, not age itself. Thus, age is the same as the current year minus the year of birth. Besides that, the change on age is a natural change, not a change caused by the person.

The second criterion was needed so that the resulting model would not be too complex. If two demographic data with quite a lot of categories are combined, the resulting model will be too complex. The number of categories for each demographic data is as follows:

1. Age group: 3 categories
2. Gender: 2 categories
3. Ethnicity: 7 categories
4. Blood type: 4 categories
5. Zodiac group: 2 categories
6. Zodiac component: 4 categories

The data above shows that ethnicity does not meet this criterion because it has too many categories. For this reason, the demographic data that meet the second criteria are age group, gender, blood type, zodiac sign, and zodiac component.

The last criterion determined was fairly good accuracy. In this criterion, the accuracy of age group - gender was among the best in all traits. Accuracy scores for age group - gender were 60.5% for extraversion, 86.8% for agreeableness, 85.0% for conscientiousness, 68.1% for emotional stability, and 54.3% for intellect. This accuracy value was similar to the prediction accuracy value using the personality elicitation from text method, namely: (1) 58% average accuracy (Argamon et al, 2005 in (Oberlander & Nowson, 2006)), (2) 56% accuracy for extraversion, agreeableness, and conscientiousness, 58% for emotional stability, and 63% for openness (Mairesse et al., 2007), (3) 63.1% average accuracy (Celli, 2012), (4) 100% accuracy for extraversion and agreeableness, 80% for conscientiousness, and 50% for neuroticism and openness (Di Rienzo & Neishabouri, 2016). Meanwhile the accuracy of the model based on blood type and zodiac sign and their combination, was not too different. Therefore, it is not yet possible to determine one model as the working model.

As it has not been possible to find one model from the three criteria above, additional criteria were used to obtain the working model. Researchers have found a very close relationship between age - gender and personality traits. Researchers such as (Caspi & Roberts, 2001), (Caspi et al., 2005), and (Hampson & Goldberg, 2006) believe that personality traits change with age. Indeed, this change is not continuous as there is a period of stability in a certain period of time. Researchers have different opinions regarding when the stable period occurs. According to (Hampson & Goldberg, 2006), this period occurs between childhood and early adulthood. (Leon et al., 1979) believe that it can also occur between middle age and old age over a period of about 30 years. Meanwhile, according to (Mõttus et al., 2012) the stable period occurs during a short period of time in old age. Although there could be a stable period at certain times in a person's life, (Harris et al., 2016) found that human personality traits always undergo small changes over time. As a result, childhood personality traits are very different from those in old age. The key periods in this change are late childhood and adolescence because it is during these periods that the influence of the gender factor is first discovered (Soto et al., 2011). Gender affects the differences in the personality of men and women in all traits. It was found that women are higher in extraversion, agreeableness, conscientiousness, and neuroticism. Meanwhile men are only higher than women in openness (Soto et al., 2011).

With regard to the relationship between blood type and personality traits, scientists basically reject the theory. They regard this as a superstitious belief or a pseudoscience ((Yamaguchi, 2005);(Buerk, 2010)). The reason is because the research results do not show a relationship between the two ((Cramer & Imaike, 2002); (Rogers & Glendon, 2003); (Wu et al., 2005); (Nawata, 2014)). In addition, prediction of personality traits based on blood type is only popular in Japan.

Meanwhile in terms of zodiac signs, recent research has also found no relationship between people's zodiac signs and their personality traits. If there are still people who believe in horoscopes and feel that the horoscope predictions are correct, this is due to the so-called Barnum effect (Azucena et al., n.d.). From this explanation, blood type and zodiac sign are considered unsuitable to be used as models. Therefore, the authors have chosen a personality trait model based on age group and gender. The resulting model based on age group and gender is presented in table 1. The more detailed explanations on how to get the personality model can be found in (Paryudi et al., 2021).

Table 1. 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

3.2. Preference Modeling

As has been explained previously that we choose three favorites (favorite 1, favorite 2, favorite 3) together with their favorite levels for each participant (an example of such data is provided in table 2).

Before the modeling step, we put together favorites in the same fashion style into one. It can be seen in table 2 that the Natural fashion style (yellow cell) is liked by participants either as favorite 1, favorite 2, or favorite 3. We put them into one and the result is shown in table 3. The data show participants' preferences on the Natural fashion style. The same process is carried out on other fashion styles: Classic, Creative, Dramatic, Elegant Chic, Feminine, and Rebellious.

Table 2. Example of three favorites on fashion together with their levels.

Favorite 1	Level	Favorite 2	Level	Favorite 3	Level
Elegant Chic	Strong	Classic	Strong	Feminine	Moderate
Classic	Moderate	Elegant Chic	Weak	Natural	Weak
Natural	Strong	Feminine	Strong	Elegant Chic	Strong
Classic	Moderate	Elegant Chic	Moderate	Natural	Moderate
Elegant Chic	Moderate	Feminine	Moderate	Natural	Moderate
Rebellious	Strong	Elegant Chic	Moderate	Feminine	Weak
Natural	Strong	Elegant Chic	Strong	Feminine	Moderate
Elegant Chic	Strong	Feminine	Strong	Natural	Moderate
Creative	Weak	Elegant Chic	Weak	Natural	Weak
Feminine	Moderate	Natural	Moderate	Classic	Moderate
Feminine	Moderate	Classic	Weak	Elegant Chic	Weak
Feminine	Weak	Elegant Chic	Weak	Natural	Weak
Natural	Moderate	Feminine	Moderate	Elegant Chic	Moderate
Classic	Weak	Elegant Chic	Weak	Feminine	Weak
Natural	Strong	Elegant Chic	Moderate	Feminine	Moderate

Table 3. Example of favorite data (on the Natural fashion style) that is used in the modeling.

Favorite	Level
Elegant Chic	Strong
Natural	Weak
Natural	Strong
Natural	Moderate
Natural	Moderate
Rebellious	Strong
Natural	Strong
Natural	Moderate
Natural	Weak
Natural	Moderate
Feminine	Moderate
Natural	Weak
Natural	Moderate
Classic	Weak
Natural	Strong

In this modeling, we make use of the association rule method, more exactly class association rule. The reason for the choice is that we want the preference on fashion is always as the consequent or class. Meanwhile, the antecedent is personality level.

To conform to the personality model shown in table 1, we only use data having low extraversion levels, high agreeableness levels, and high conscientiousness levels. Besides that, we only use favorite data with moderate and strong levels. The separation between male and female must be done in this modeling since it is not possible to recommend male fashion to females and vice versa.

The result of the modeling can be seen in figure 1. The figure exhibits the relationship between personality and preference on fashion. Note that the numbers on the cells are the confidence of the rules. When this figure is combined with the personality model shown in table 1, the result is shown in figure 2.

To determine the favorite fashion style of each personality, we pick two fashion styles with the highest confidence (yellow cell in Figure 2). Based on this, the preference model can also be presented differently as shown in Figure 3.

These two models, personality and favorite models will be applied in a recommender system whose architecture is shown in Figure 4. The system starts when an active user enters his/her year of birth and gender. Based on those data, using Model 1 (personality model), the system classifies his/her personality. This personality data is then sent to Model 2 (preference model) to classify his/her preference based on his/her personality. After getting the user's preference, the system then fetches all neighbors in the user database to find neighbors with the same year of birth, gender, and preference as the year of birth, gender, and preference of the active user. The result of this activity is a list of neighbors having the same year of birth, gender, and preference. Using this list, the system then finds the nearest neighbor. Once the nearest neighbor is obtained, the system takes all items that have been consumed by the nearest neighbor. These are the items that will be recommended to the active user. After consuming some or all the recommended items, the active user data are saved into the user database.

Figure 1. Modeling result relating personality E(xtraversion), A(greeableness), C(onscientiousness), E(motional) S(tability), and I(ntellect) with fashion style Cl(assic), Cr(eative), Dr(amatic), E(legant) C(hic), Fe(minine), Na(tural), dan Re(bellious).

	Personality Traits					Fashion Style						
	E	**A**	**C**	**ES**	**I**	**Cl**	**Cr**	**Dr**	**EC**	**Fe**	**Na**	**Re**
Male	Low	High	High	Low	Low	0,80	0,50	0,45	0,60	0,50	0,60	0,50
	Low	High	High	Low	High	0,58	0,46	0,46	0,43	0,70	0,80	0,38
	Low	High	High	High	Low	0,50	0,67	0,67	0,80	0,50	0,80	0,67
	Low	High	High	High	High	0,56	0.26	0,26	0,67	0,59	0,94	0,28
Female	Low	High	High	Low	Low	0,58	0,35	0,38	0,97	0,33	0,86	0,35
	Low	High	High	Low	High	0,39	0,42	0,39	0,94	0,34	0,88	0,43
	Low	High	High	High	Low	0,45	0,34	0,36	1,00	0,32	0,81	0,34
	Low	High	High	High	High	0,47	0,53	0,50	0,94	0,47	0,94	0,55

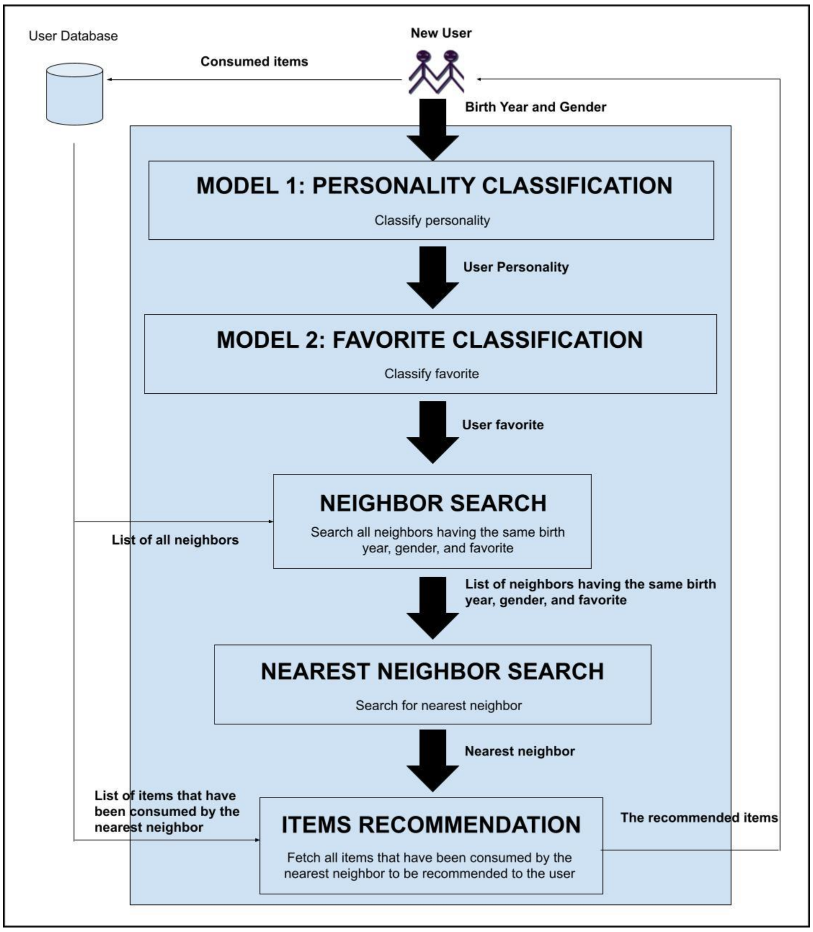
Figure 2. Merging between personality model (table 1) and preference modeling result (figure 1). Yellow cells indicate the favorite fashion style of each personality trait (E: Extraversion, A: Agreeableness, C: Conscientiousness, ES: Emotional Stability, I: Intellect, Cl: Classic, Cr: Creative, Dr: Dramatic, EC: Elegant Chic, Fe: Feminine, Na: Natural, Re: Rebellious).

	Personality Traits					Fashion Style						
	E	A	C	ES	I	Cl	Cr	Dr	EC	Fe	Na	Re
Male	Low	High	High	Low	High	0,58	0,46	0,46	0,43	0,70	0,80	0,38
	Low	High	High	High	Low	0,50	0,67	0,67	0,80	0,50	0,80	0,67
	Low	High	High	High	High	0,56	0.26	0,26	0,67	0,59	0,94	0,28
Female	Low	High	High	Low	Low	0,58	0,35	0,38	0,97	0,33	0,86	0,35
	Low	High	High	High	Low	0,45	0,34	0,36	1,00	0,32	0,81	0,34
	Low	High	High	High	Low	0,45	0,34	0,36	1,00	0,32	0,81	0,34

Figure 3. The preference model.

Gender	Personality Traits					Favorite Fashion Style
Male	Low Extraversion	High Agreeableness	High Conscientiousness	Low Emotional Stability	High Intellect	Natural, Feminine
	Low Extraversion	High Agreeableness	High Conscientiousness	High Emotional Stability	Low Intellect	Elegant Chic, Natural
	Low Extraversion	High Agreeableness	High Conscientiousness	High Emotional Stability	High Intellect	Natural, Elegant Chic
Female	Low Extraversion	High Agreeableness	High Conscientiousness	Low Emotional Stability	Low Intellect	Elegant Chic, Natural
	Low Extraversion	High Agreeableness	High Conscientiousness	High Emotional Stability	Low Intellect	Elegant Chic, Natural
	Low Extraversion	High Agreeableness	High Conscientiousness	High Emotional Stability	Low Intellect	Elegant Chic, Natural

Figure 4. The architecture of the proposed recommender system.



4. Conclusion

Two models, personality and preference models have been created. The personality model relates demographic data and the big five-based personality traits, namely agreeableness, emotional stability, intellect, extraversion, conscientiousness. Meanwhile preference model correlates personality traits and preference on fashion style. There are seven predominant fashion styles: classic, creative, dramatic, elegant chic, feminine, natural, and rebellious.

We create 325 models for the five traits or 65 models for each trait. Three criteria are used to select the working model, namely the demographic data does not change entire life, not too many categories in each demographic data, has fairly high accuracy. Based on those criteria, we pick a model that is based on age group and gender as the working model since the model fulfills all criteria. Another reason to choose the model is that previous research finds that there is a very close relationship between ages – gender and personality traits.

The working model, which is based on age group and gender, shows low extraversion, high agreeableness, and high conscientiousness. Meanwhile, for emotional stability and intellect, the personality levels vary depending on the age group – gender cohort: (1) Adolescence male has low emotional stability and high intellect. (2) Adolescence female has low emotional stability and low intellect. (3) Adulthood male has high emotional stability and low intellect. (4) Adulthood female has high emotional stability and low intellect. (5) Middle age male has high emotional stability and high intellect. (6) Middle age female has high emotional stability and low intellect.

To conform to the aforementioned personality model, we only use data with low extraversion, high agreeableness, and high conscientiousness in the preference modeling. Besides that, in the modeling, we also only use data whose favorite levels are strong or moderate. From this modeling, we obtain the following results: (1) Male with low emotional stability and high intellect likes Natural and Feminine fashion style. (2) Male with high emotional stability and low intellect likes Elegant Chic and Natural fashion style. (3) Male with high emotional stability and high intellect likes Natural and Elegant Chic fashion style. (4) Female with low emotional stability and low intellect likes Elegant Chic and Natural fashion style. (5) Female with high emotional stability and low intellect like Elegant Chic and Natural fashion style.

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

Azucena, A. A. G., Balaba, B. I. S., & Paez, E. J. L. (n.d.). *Psychology of Astrology (The relationship between zodiac signs and the personality of an individual)*.

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

Buerk, R. (2010). *Dating by blood type in Japan*. http://news.bbc.co.uk/2/hi/8646236.stm

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

Caspi, A., & Roberts, B. W. (2001). Personality Development Across the Life Course : The Argument for Change and Continuity. *Psychological Inquiry*, *12*(2), 49–66. https://doi.org/10.1207/S15327965PLI1202

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality Development: Stability and Change. *Annual Review of Psychology*, *56*(1), 453–484. https://doi.org/10.1146/annurev.psych.55.090902.141913

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

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

Chen, L., Wu, W., & He, L. (2013). How Personality Influences Users' Needs for Recommendation Diversity? *CHI 2013: Changing Perspective*, 829–834. http://movie.douban.com/

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

Cramer, K. M., & Imaike, E. (2002). Personality, blood type, and the five-factor model. *Personality and Individual Differences*, *32*(4), 621–626. https://doi.org/https://doi.org/10.1016/S0191-8869(01)00064-2

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

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

Hampson, S. E., & Goldberg, L. R. (2006). A first large cohort study of personality trait stability over the 40 years between elementary school and midlife. *Journal of Personality and Social Psychology*, *91*(4), 763–779. https://doi.org/10.1037/0022-3514.91.4.763

Harris, M. A., Brett, C. E., Johnson, W., & Deary, I. J. (2016). Personality Stability From Age 14 to Age 77 Years. *Psychology and Aging*, *31*(8), 862–874. https://doi.org/10.1037/pag0000133.supp

Heller, D., Perunovic, W. Q. E., & Reichman, D. (2009). The future of person-situation integration in the interface between traits and goals: A bottom-up framework. *Journal of Research in Personality*, *43*(2), 171–178. https://doi.org/10.1016/j.jrp.2008.12.011

Hu, R. (2010). Design and User Issues in Personality-based Recommender Systems. *RecSys'10*, 386.

Kraaykamp, G., & van Eijck, K. (2005). Personality, media preferences, and cultural participation. *Personality and Individual Differences*, *38*(7), 1675–1688. https://doi.org/10.1016/j.paid.2004.11.002

Leon, G. R., Gillum, B., Gillum, R., & Gouze, M. (1979). Personality Stability and Change Over a 30-Year Period--Middle Age to Old Age. *Journal of Consulting and Clinical Psychology*, *47*(3), 517–524. https://doi.org/10.1037/0022-006X.47.3.517

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

Maxfield, C. (2017). *7 Predominant Personality Styles*. https://claremaxfield.com.au/blog/men/

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., Hřebíčková, 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*.

Mõttus, R., Johnson, W., & Deary, I. J. (2012). Personality Traits in Old Age: Measurement and Rank-Order Stability and Some Mean-Level Change. *Psychology and Aging*, *27*(1), 243–249. https://doi.org/10.1037/a0023690

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.

Nawata, K. (2014). No relationship between blood type and personality: evidence from large-scale surveys in Japan and the US. *Shinrigaku Kenkyu: The Japanese Journal of Psychology*, *85*(2), 148–156. https://doi.org/10.4992/jjpsy.85.13016

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

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the Prediction of Consequential Outcomes. *Annual Review of Psychology*, *57*(1), 401–421. https://doi.org/10.1146/annurev.psych.57.102904.190127

Paryudi, I., Ashari, A., & Mustofa, K. (2021). Modelling Relationship between Demographic Data and Personality Traits. *Turkish Journal of Computer and Matemathics Education*, *12*(14), 2247–2255.

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

Rentfrow, P. J., Goldberg, L. R., & Zilca, R. (2011). Listening, Watching, and Reading: The Structure and Correlates of Entertainment Preferences. *Journal of Personality*, *79*(2), 223–258. https://doi.org/10.1111/j.1467-6494.2010.00662.x

Rentfrow, P. J., & Gosling, S. D. (2003). The Do Re Mi's of Everyday Life: The Structure and Personality Correlates of Music Preferences. *Journal of Personality and Social Psychology*, *84*(6), 1236–1256. https://doi.org/10.1037/0022-3514.84.6.1236

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

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
Wu, K., Lindsted, K. D., & Lee, J. W. (2005). Blood type and the five factors of personality in Asia. *Personality and Individual Differences*, *38*(4), 797–808. https://doi.org/https://doi.org/10.1016/j.paid.2004.06.004
Yamaguchi, M. (2005). *Myth about Japan blood types under attack*. https://web.archive.org/web/20091228182320/http:/aol.mediresource.com/channel_health_news_details.asp?news_id=6661&news_channel_id=11&channel_id=11