Application of Decision Tree in Big Five Personality Traits

Rhea Elka Pandumpi 13519047
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13519047@std.stei.itb.ac.id

Abstract—**Discovering depths to a personality has become a trend to expect a more accountable approach to life choices such as careers and building relationships. One of the most famous personality tests used to measure a person's traits is "Big Five". This paper discusses about the usage of Decision Tree in the personality test and how the result of the test can actually be applied in daily life.**

Keywords—**big five, personality traits, decision tree, career choice.**

I. INTRODUCTION

Nowadays, people find that discovering their personality is something that is important. The importance of discovering one's personality is to make more authoritative life choices, such as career, vocation, or maintaining relationship with others. Personality tests have been something that is favourable in popular culture, because people believe they can know themselves even better.

Some of the most famous personality tests include Myers-Briggs and Big Five tests. However, many personality psychologists consider Big Five tests superior to the Myers-Briggs Type Indicator. That is primarily because the Big Five tests were developed using the scientific method. The types in Myers-Briggs Type Indicator categorize people as extremes, when human qualities are better represented by a continuum.

Big Five is built on five dimensions or clusters of personality variances in the 1930s by American psychologists Gordon Allport and S.Odbert. To make knowing oneself easier, Big Five Personality Traits broke down the differences into five: Openness to experience, Conscientiousness, Extroversion, Agreeableness, and Neuroticism.

By knowing how much someone has the five traits, applying leaderships in teams, choosing career choices, and making life decisions are expected to be easier and more accurate. Decision tree can be used to determine which traits are high or low in someone. The nodes of the decision tree will be used to place the questions that are related to the five traits.

II. THEOREMS

1. Tree
   A tree is a collection of nodes connected by directed (or undirected) edges. A tree is a *nonlinear* data structure, compared to arrays, linked lists, stacks and queues which are linear data structures. A tree can be empty with no nodes or a tree is a structure consisting of one node called the root and zero or one or more subtrees. A tree has following general properties:
   - One node is distinguished as a root;
   - Every node (exclude a root) is connected by a directed edge *from* exactly one other node; A direction is: *parent -> children.*
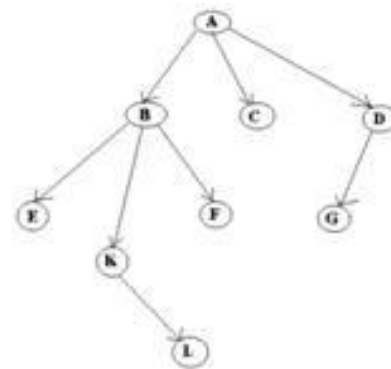

Image 1 Tree

A is a parent of B, C, D, B is called a child of A. On the other hand, B is a parent of E, F, K. Each node can have *arbitrary* number of children. Nodes with no children are called leaves, or external nodes. In the above picture, C, E, F, L, G are leaves. Nodes, which are not leaves, are called internal nodes. Internal nodes have at least one child.

Nodes with the same parent are called siblings. In the picture, B, C, D are called siblings. The depth of a node is the number of edges from the root to the node. The depth of K is 2. The height of a node is the number of edges from the node to the deepest leaf. The height of B is 2. The height of a tree is a height of a root.

2. General Tree
   A general tree is a tree where each node may have zero or more children (a binary tree is a specialized case of a general tree). General trees are used to model applications such as file systems.
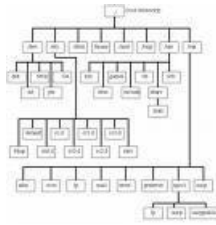
Image 2 General Tree
Figure courtesy of www.washington.edu

3. Implementation

Since each node in a tree can have an arbitrary number of children, and that number is not known in advance, the *general* tree can be implemented using a first child/next sibling method. Each node will have TWO pointers: one to the leftmost child, and one to the rightmost sibling. The following picture illustrates this.
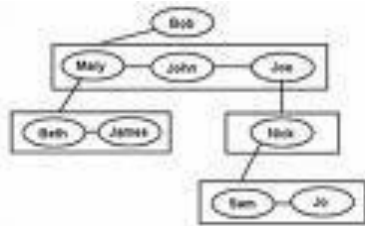

Image 3 Implementation
Figure courtesy of www.gamedev.net

4. Binary Trees

We will see that dealing with binary trees, a tree where each node can have no more than two children is a good way to understand trees. A binary tree in which each node has exactly zero or two children is called a full binary tree. In a full tree, there are no nodes with exactly one child.
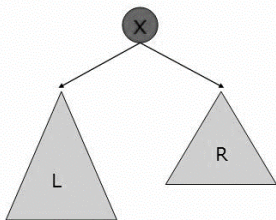

Image 4 Binary Trees

5. Decision Tree

A decision tree is a tool that builds regression models in the shape of a tree structure. Decision trees take the shape of a graph that illustrates possible outcomes of different decisions based on a variety of parameters. Decision trees break the data down into smaller and smaller subsets, they are typically used for machine learning and data mining, and are based on machine learning algorithms. Decision trees are also referred to as recursive partitioning.

Decision Tree Components

Decision trees are made up to two parts: nodes and leaves. Nodes: represent a decision test, examine a single variable and move to another node based on the outcome. Leaves: represent the outcome of the decision.

6. The Usage of Decision Tree

Decision trees are useful to make various predictions. For example, to predict if an email is SPAM or not, to predict health outcomes, to predict what group an individual belongs to based on a variety of factors that are specified in the decision tree model.

7. Advantages of Decision Tree

The advantages of decision tree are that they are simple to understand and interpret, help determine the expected outcomes of various scenarios, help determine best and worst values for different scenarios, can be combined with other decision techniques, require a relatively low degree of data preparation, can accommodate missing data, low sensitivity to outliers, low impact of nonlinear relationships between parameters, can handle both categorical and numeric variables, and can translate the decision tree results into "decision rules".

8. Disadvantages of Decision Tree

The disadvantages of decision tree are that for categorical variables, more levels of the variable creates more bias of the decision tree toward that variable, and if the tree is *over-fitted* to the data, the results can be poor predictors.

9. Big Five

The Big Five are five broad factors (dimensions) of personality traits. They are:

1. Extraversion (sometimes called Surgency). The broad dimension of Extraversion encompasses such more specific traits as talkative, energetic, and assertive.

   Extraversion (also "extroversion") is marked by pronounced engagement with the external world. Extraverts enjoy being with people, are full of energy, and often experience positive emotions. They tend to be enthusiastic, action-oriented individuals who are likely to say "Yes!" or "Let's go!" to opportunities for excitement. In groups they like to talk, assert themselves, and draw attention to themselves.

   Introverts lack the exuberance, energy, and activity levels of extraverts. They tend to be quiet, low-key, deliberate, and less dependent on the social world. Their lack of social involvement should not be interpreted as shyness or depression; the introvert simply needs less stimulation than an extravert and more time alone to re-charge their batteries.

2. Agreeableness. Includes traits like sympathetic, kind, and affectionate.

   Agreeableness reflects individual differences in concern with cooperation and social harmony. Agreeable individuals value getting along with others. They are therefore considerate, friendly, generous, helpful, and willing to compromise their interests with others'. Agreeable people also have an optimistic view of human nature. They believe people are basically honest, decent, and trustworthy.

Disagreeable individuals place self-interest above getting along with others. They are generally unconcerned with others' well-being, and therefore are unlikely to extend themselves for other people. Sometimes their skepticism about others' motives causes them to be suspicious, unfriendly, and uncooperative.

Agreeableness is obviously advantageous for attaining and maintaining popularity. Agreeable people are better liked than disagreeable people. On the other hand, agreeableness is not useful in situations that require tough or absolute objective decisions. Disagreeable people can make excellent scientists, critics, or soldiers.

3. Conscientiousness. Includes traits like organized, thorough, and planful.

Conscientiousness concerns the way in which we control, regulate, and direct our impulses. Impulses are not inherently bad; occasionally time constraints require a snap decision, and acting on our first impulse can be an effective response. Also, in times of play rather than work, acting spontaneously and impulsively can be fun. Impulsive individuals can be seen by others as colorful, fun-to-be-with, and zany.

Conscientious individuals avoid trouble and achieve high levels of success through purposeful planning and persistence. They are also positively regarded by others as intelligent and reliable. On the negative side, they can be compulsive perfectionists and workaholics. Furthermore, extremely conscientious individuals might be regarded as stuffy and boring. Unconscientious people may be criticized for their unreliability, lack of ambition, and failure to stay within the lines, but they will experience many short-lived pleasures and they will never be called stuffy (i.e. dull, boring, unimaginative).

4. Neuroticism (sometimes reversed and called Emotional Stability). Includes traits like tense, moody, and anxious.

Neuroticism, also known inversely as *Emotional Stability*, refers to the tendency to experience negative emotions. Those who score high on Neuroticism may experience primarily one specific negative feeling such as anxiety, anger, or depression, but are likely to experience several of these emotions. People high in Neuroticism are emotionally reactive. They respond emotionally to events that would not affect most people, and their reactions tend to be more intense than normal. They are more likely to interpret ordinary situations as threatening, and minor frustrations as hopelessly difficult.

Their negative emotional reactions tend to persist for unusually long periods of time, which means they are often in a bad mood. These problems in emotional regulation can diminish a neurotic's ability to think clearly, make decisions, and cope effectively with stress.

At the other end of the scale, individuals who score low in Neuroticism are less easily upset and are less emotionally reactive. They tend to be calm, emotionally stable, and free from persistent negative feelings. Freedom from negative feelings does not mean that low scorers experience a lot of positive feelings; frequency of positive emotions is a component of the Extraversion domain.

5. Openness to Experience (sometimes called Intellect or Intellect/Imagination). Includes traits like having wide interests, and being imaginative and insightful.

Openness to Experience describes a dimension of personality that distinguishes imaginative, creative people from down-to-earth, conventional people. Open people are intellectually curious, appreciative of art, and sensitive to beauty. They tend to be, compared to closed people, more aware of their feelings. They therefore tend to hold unconventional and individualistic beliefs, although their actions may be conforming (see agreeableness).

People with low scores on openness to experience tend to have narrow, common interests. They prefer the plain, straightforward, and obvious over the complex, ambiguous, and subtle. They may regard the arts and sciences with suspicion, regarding these endeavors as abstruse or of no practical use. Closed people prefer familiarity over novelty; they are conservative and resistant to change.

III. APPLICATION OF DECISION TREE IN BIG FIVE PERSONALITY TRAITS

In Big Five Personality tests, the participants will be given a questions which they must answer based on their perception of themselves. The questions will be related to the five broad factors, namely extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. To conclude the test result, a decision tree can be used based on the answers that the participants give.

The result of the test can be used to determine the communication style and career choices.

The correlation between the qualities, communication style, and career choices can be seen in the table below.

No	Quality	Communication Style	Career Choices
1	High openness	Direct, honest, new ideas	Artists, entrepreneurs, engineers
2	Low openness	Respectful of personal beliefs	Bankers, analysts, contractors
3	High conscientiousness	Respectful of time, direct expression of thoughts	Freelancers, consultants, business owners

No			
4	Low conscientiousness	Need for spontaneity	Firefighters, mechanics, drivers,
5	High extroversion	Casual conservation	Actors, counselors, photographers
6	Low extroversion	Clear, brief	Librarians, engineers, writers
7	High agreeableness	Vulnerable and open to building personal connection	Therapists, nurses, teachers
8	Low agreeableness	Direct, logical, patient	Accountants, programmers, scientists
9	High neuroticism	Encouraging, supportive	Florists, writers, accountants
10	Low neuroticism	Open, upbeat, positive	Police officers, lawyers, surgeons

Table 1 The correlation between the Big Five Personality Traits and preferred comunication styles and career choices, based on https://www.crystalknows.com/big-five

Using the table above, not only that people can know themselves better, but they can also know how to communicate better or what communication style to choose when conversing with people with those traits. For example, to encounter a person with high openness, low conscientiousness, high agreeableness, and high neuroticism, it is better to balance the communication styles that are preferred by those traits. With this example, we would need to communicate directly, honestly, but also leaving space for spontaneity, open to build personal connection and also supportive to the person.

For the career choice, one can have more considerations because of the many traits they possess (not only in types). The advantage of Big Five personality traits are that the traits are independent of each other, so even though one of someone's traits might not be so compatible with the career, but they can know better whether any of their other traits can complement the career despite of the lacking traits.

Besides of the general career choices given above, we can take one example field where different qualities can actually work. In this example, we will take the psychology field to show the role of the different qualities.

No	Qualities	Career Choice
1	High openness	Psychology researcher

No	Qualities	Career Choice
2	Low openness	Psychometrist
3	High conscientiousness	Bioethicist
4	High easy-goingness	Art, music, horticultural, or pet therapist
5	High extroversion	Salesperson/psychology related products
6	Low extroversion	Lab technician/technologist
7	High agreeableness	Human resources specialist
8	Low agreeableness	Self-employed consultant
9	High neuroticism	Therapist in solo private practice, career counselor, personal coach
10	Low neuroticism	Hostage negiator/psychologist in a mental hospital

Table 2 The compatibility between the Big Five Personality Traits with jobs in psychology field, based on https://www.psychologytoday.com/us/blog/how-do-life/201908/big-five-careers

Observing the table above, it is very likely that different traits can be useful in one field, with different kinds of specialties. For example, a person with high openness can choose to be a psychology researcher, but a person that has lower openness can still have the chance to a compatible career in psychology, namely bioethicist.

To decide whether a person is high or low in those traits, a decision tree can be used, with every node is a question. On the root, the question given is: "Do you enjoy abstract things?". This is used to determine openness. The "yes" answer is represented by the blue edge, and the "no" answer is represented by the red edge. The "yes" will then lead the final answer to high openness, and contrarily, the "no" will lead to low openness. The next question is to determine conscientiousness, "Do you hate to change your schedules last minute?".

The "yes" answer will mean high in said trait, and so on with the next questions: "Do you prefer more social interactions?", "Can you emphatize with people's feelings?", with "yes" answers will mean that the person is high in extroversion and agreeableness, sequentially, and the "no" answers will mean lowness in the trait. The only exception is when the answer is "yes" to the last question: "Are you comfortable with yourself?". The "yes" will mean lowness in neuroticism and the "no" will mean highness in neuroticism. In the end of the node, the result will show whether someone is high or low in the five traits. For example, if the answer is "yes" to all the questions, it will be displayed that the person tested has high OCEA (Opennesss, Conscientiousness, Extroversion, and Agreeableness) and low Neuroticism. And conversely, if the answers are all "no", the results will be contrast with the one before.
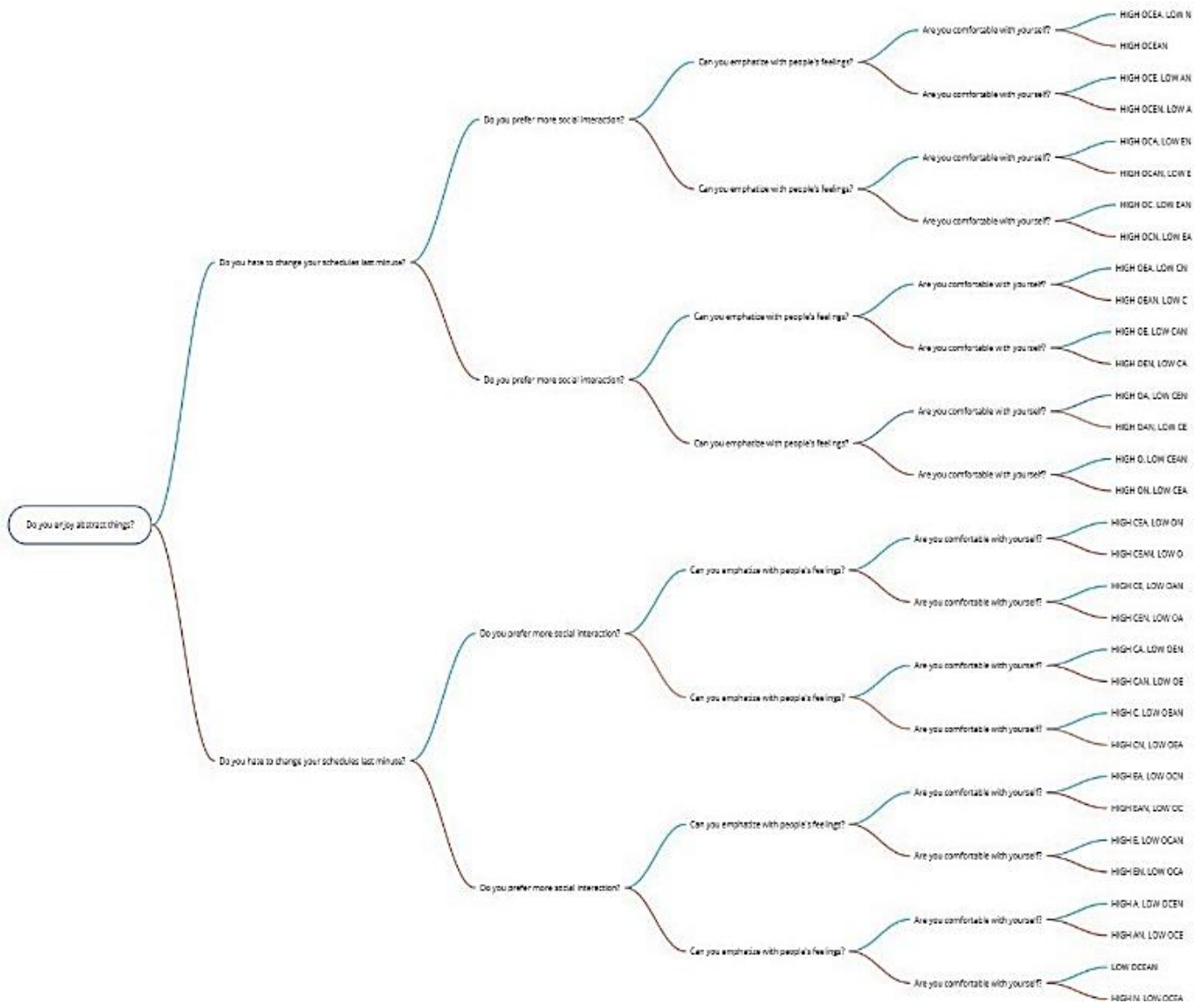
Image 5 Visualization of Decision Tree

One of the disadvantages of the decision tree in this approach is that it is quick to assume that a person is "high" or "low" in said traits when there is moderation when measuring those traits. However, adequate accuracy might be secured because when there is only two answers, when one answer is being chosen, the answer is most likely to be perceived as truer to one's own personality. Most career choices and relationship-building advices are more directed to "low" or "high" in traits, the moderate in those traits are usually ignored.

IV. CONCLUSION

Decision tree can be used to determine a quick result to OCEAN (Big Five) personality traits with using questions that are related with the five traits to measure each of the traits in a person. The results can be used to more accountable approach to career choices and building relationships (such as communication styles). Due to the nature of the independence of each trait, the traits that scored high are expected to complement the ones that scored low, so the career choices are not always "black-and-white" to people that might have

can also be used to show that different scores in different traits can work in one same field.

V. ACKNOWLEDGMENT

Thanks be to God for being there for me, for abundantly bestowing His graces upon me from the moment I started this paper to the finish. I also want to thank my family, close friends who would always support me in my every way, and everyone who always motivates me to excel myself everyday. Last, but not least, a huge gratitude is due to my lecturer, Mrs. Fariska Zakhralativa Ruskanda, S.T., M.T. for her wonderful guidance in Discrete Mathematics in the third semester.

REFERENCES

[1] https://www.psychologytoday.com/us/basics/big-5-personality-traits. Accessed on December 9, 2020.
[2] https://www.psychologytoday.com/us/blog/people-are-strange/201910/how-accurately-measure-your-personality. Accessed on December 9, 2020.
[3] https://www.michiganstateuniversityonline.com/resources/leadership/lead-your-team-with-big-five-model. Accessed on December 9, 2020.

[4] http://www.cs.cmu.edu/~clo/www/CMU/DataStructures/Lessons/lesson4_1.htm. Accessed on December 9, 2020.

[5] https://ademos.people.uic.edu/Chapter24.html. Accessed on December 9, 2020.

[6] https://www.psychologytoday.com/us/blog/how-do-life/201908/big-five-careers. Accessed on December 10, 2020.

[7] https://www.crystalknows.com/big-five. Accessed on December 10, 2020.

[8] https://library.cityvision.edu/big-five-personality-traits. Accessed on December 10, 2020.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 10 Desember 2020

Rhea Elka Pandumpi 13519047