

# Knowing Personality Traits on Facebook Status Using the Naïve Bayes Classifier

Mohammad Zoqi Sarwani<sup>1</sup>, Muhammad Shubkhan Salafudin<sup>2</sup>, Dian Ahkam Sani<sup>3</sup>

<sup>1,2,3</sup>Universitas Merdeka Pasuruan, Pasuruan, 67129, Indonesia

<sup>1</sup>zaki\_swr@yahoo.com (zoqi.sarwani@unmerpas.ac.id)\*; <sup>2</sup>shubkhansalafudin@gmail.com; <sup>3</sup>achcomes@gmail.com

\*Corresponding Author

---

## ABSTRACT

---

With the development of social media trends among students by using Facebook social media, students can communicate and pour out everything that is felt in the form of status. Personality is the character or various characters of a person - therefore, how a person to adjust to the surrounding environment for the achievement of communication smoothly. In the personality category, many things classify a person's category in the psychologist theory. In this exercise, the Big Five, the psychologist theory, is described in five codes, namely Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism. Naive Bayes Classifier is used to determine the highest probability value with the aim to determine the highest value. The data used are two namely training data and testing data obtained from the Facebook status of students. From the data obtained can be tested in the system that the accuracy value is 88%.

Keywords: Facebook status; Knowing Personality Traits; *Big five*; Naive Bayes Classifier

---

### Article History

Received : June, 02<sup>nd</sup> 2020

Revised : June, 18<sup>th</sup> 2020

Accepted : June, 30<sup>th</sup> 2020

---

## I. INTRODUCTION

Communication is done in meeting human needs to interact with others related to human function as social beings. Almost all people in the era like now social networking communication are widely used by using tools from social media. Social media is composed of two words, namely "media" and "social". "Media" is meant as a communication tool. [1] Many social media have become phenomenal in society, namely Facebook, Twitter, YouTube, Instagram, and Path, which are various types of social media that are in great demand, especially among students.

When needs are done, there will be various things that influence them. Nowadays social media is widely used, because it invites anyone who is quickly attracted to participate by giving feedback openly, commenting, and sharing information in a fast and unlimited time it can be ensured that each individual has a smartphone, also has social media accounts, such as *Facebook, Twitter, Path, Instagram, and so on*. This condition is already prevalent, which changes the pattern of life how to communicate in the current digital era.

Personality is abstract and very broad so that science related to humanity uses the term Personality or psyche is the scope of the whole mind, feeling, behavior, awareness, and unconsciousness. Personality is a unique and unique aspect of a person's behavior, which can affect a person's ability to adapt to the environment [2].

Psychologists study several tests to be able to define a person's character. It seems clear in front of other people is usually a habit, and from this habit, we can find a person's character traits [3]. Personality is abstract and very broad so that science related to humanity uses the term Personality or psyche is the scope of the whole mind, feeling, behavior, consciousness, and unconsciousness. Personality is a unique and unique aspect of a person's behavior, which can affect a person's ability to adapt to the environment. With several theories that can be used are usually used in the process of testing a psychologist, namely: MBTI (*Myers Briggs Type Indicator*) to measure personality tests [4] Each IST (*Intelligence Standart Test*) takes 60 minutes, in this case, if using the IST MBTI test only takes 20 minutes and Kraepelin 12.5 minutes [4]. DISC (*Dominance, Influence Steadiness, Compliance*) is a personality measurement tool used to view and measure work behavior based on four types. In using the ITS DISC test, the company uses a sheet in which a graphic is filled out by participants (*employees*). DISC can only be used in decision support systems because the results obtained from this system are a weight value used to place employees in a part of the company [4].

Big Five Personality is one of the theories of personality that developed in the 20th century. Big five, namely Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness, including constructs that can be measured from several entities in the form of behavior. With personality formed from biological behavior and learning is seen as an important mental and social factor that forms a person's unique responses to environmental and static stimuli over time and situations [5].

The need to determine the source of knowledge in the document to trigger in the form of text is used text mining in using the application of processes in text mining, and it will obtain data patterns, trends, and extraction of potential knowledge from text data. Naïve Bayes is a simple probabilistic classification that calculates a set of probabilities by adding up the frequency and combination of values from a given dataset. Another definition says Naïve Bayes is a classification by probability method by

looking at Twitter analysis research to find out the character of someone using the Naïve Bayes Classifier. The results of this study are the development of previous research, which uses the Probabilistic Neural Network algorithm to classify one's personality through Facebook social media [6].

In this study, the author is more or less inspired by the references of previous studies by looking at the background of the problems in this study. The research related to this research are Big Five Personality Profile and Collecting Behavior of LPs Collectors [7] Collectors generally have trait openness to high experience personalities, collectors tend their hearts and persistence in caring for hunting and more. Collectors easily join and socialize with others because collecting related to music can eliminate anxiety. Furthermore, researchers use Twitter Analysis Reference to Know Someone's Character Using the Naïve Bayes Classifier Algorithm [8] Proof of the study using MBTI psychologist tests by looking at tweets or posts on Twitter can analyze someone's personality with accurate results. From using the Naïve Bayes classifier probability algorithm, the results of the test can get a good level of accuracy. Implementation of Thesis Text Mining Classification Using the Naïve Bayes Classifier Method [9], looking at this research in the Naïve Bayes Classifier algorithm can obtain accurate data classification values which are proven in thesis testing in the library of informatics engineering study program with three categories, namely software engineering, network-based computing, and visual intelligent computing with 90% accuracy. Next sentiment analysis of category classification of public figures on Twitter [10].

The Influence of Social Media on the Students' Personality STEBIS IGM Palembang [11], this study explains that technological developments in the era of globalization are very influential on students, especially on social media which cannot be denied any more because social media invites users to participate openly, social media also has an influence to users in terms of clothing to the personality of the students of IGM, because of this modern development added that the one factor that shapes the personality itself is culture.

## II. METHOD

The flowchart of the system planning in this study with the following steps must be taken in Fig.1. The system design flow above explains the network system, knowing the personality obtained from two data, namely training data and testing data. In the training data explained, starting from taking a document (Facebook) that is used is a different ID. After getting the data in the form of text, it will in the process of using text mining, which is processed with the stages of *text processing*, *feature selection*, *stemming*. After the data is processed with *text processing*, it will get the data in standard text / standard words. To be input into the database In the testing data flow explained the same as the process of taking data from Facebook and carried out the same process to get data in the form of raw words Data - data that has been inputted into the database will be analyzed by the naïve method *Bayes classifier* for getting results.

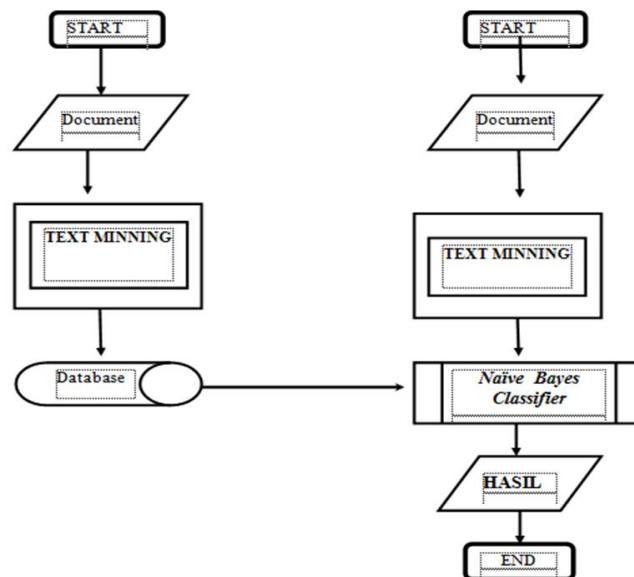


Fig 1. System Flow

### A. Personality

Formerly the mask was used in the theatrical show to show the characters that were played with that personality had meanings like a mask. Personality is an explanation of character because the character of people intends to impose norms on the person being discussed, in this case, it is said that the attitude. Attitude and behavior are viewed in terms of norms. As explained by Allport that "Character is personality evaluated and personal character is devaluated" (Allport, 1973, p52) that character and personality are the same. Another theory of psychoanalysis pioneered by Sigmund Freud views personality as consisting of three components, namely id (instinct), ego (consciousness or 'I'), and superego (conscience). The interaction between the three components is manifested in behavior

Personality is a unique and unique aspect of a person's behavior, which can affect a person's ability to adapt to the environment (Adisti, 2010, p.19). Personality itself includes thought patterns, feelings, and behavior, which is one of the unique things in each person and is a character that distinguishes one another. Based on the experts above it can be understood that the definition of personality, in general, is the characteristics and tendencies of the traits or characteristics possessed by individuals in looking at themselves and interacting with others formed by heredity and external factors such as the environment, social, and culture [1].

#### B. Facebook

Facebook is a social networking website where users can join communities such as cities, work, schools, and regions to connect and interact with others. Facebook social networking site penetrates 100 million active users announced by Facebook CEO. Facebook is currently one of the number one sites in the world until now included in the ranks of the top five most known for having many members. Entering 2006, the use of Friendster and MySpace began to be displaced by Facebook. This site with a more modern appearance allows people to get acquainted and access information to the widest possible extent. Facebook is a free internet social networking service where we can form a network by inviting our friends.

#### C. Big Five

The theory of personality used by researchers in this study is the Big Five Personality model is a theory of personality models that is determined based on five basic factors that underlie each other and include the most significant variations in human personality. Big Five is one of the methods known in the world of psychology to interpret one's personality, especially to find the relationship of personality with the work environment. The Big Five personality consists of openness (O), conscientiousness (C), extraversion (E), Agreeableness (A), and neuroticism (N) [5].

##### 1. Extraversion

*Extraversion* describes the character of someone who tends to be affectionate, cheerful, happy to talk, happy to gather, fun (McCrae and 14 Costa, in Feist & Feist, 2010) passionate, enthusiastic, dominant, friendly and communicative. People with *Extraversion* personalities are also passionate, friendly, and communicative. Someone with a high level of extraversion can be faster to make friends than someone with a low level of extraversion. Extraversion is easily motivated by change, variation in life, and also easily bored, while someone with a low level of extraversion tends to be calm and withdraw from their surroundings.

##### 2. Agreeableness or Agreement

*Agreeableness* can be characterized as being able to adapt socially well. Individuals who have a high level of Agreeableness include those who are soft-hearted, kind, and warm, tend to trust people easily and are friendly, like to help others, forgiving and direct to the problem. This personality distinguishes between soft-hearted individuals who do not know compassion. Someone with low Agreeableness likes to be suspicious, miserly, unfriendly, easily offended, tends to be more aggressive, and criticizes others and is less cooperative. Someone with agreeableness personality is also called social adaptability or likability, which means to characterize someone friendly, has a personality that always succumbs, and avoids conflicts.

##### 3. Conscientiousness

*Conscientiousness* describes someone who tends to be organized, cautious (Friedman & Schustack, 2006; McCrae and Costa, in Feist & Feist, 2010), reliable, responsible, hardworking, on time, and able to survive. Personality conscientiousness is usually referred to as dependability, impulse control, and will to achieve. Someone with a high level of conscientiousness is a hardworking, careful, punctual, and determined person. Whereas someone with a low level of conscientiousness tends to be disorganized, neglectful, lazy, and lacks purpose and gives up easily when encountering difficulties.

##### 4. Neuroticism or Neuroticism

*Neuroticism* describes someone who tends to be nervous, sensitive, tense, temperamental anxious, self-loving, very self-aware, emotional and prone to stress-related disorders

##### 5. Openness to Experience or Openness

*Openness* to experience describes someone who tends to look imaginative, creative, fun, artistic, full of curiosity, open, and prefers variation. Openness describes the complexity or breadth of life. Individuals who have high openness are generally creative, imaginative, curious, and have broad interests. As well as fun and artistic, openness distinguishes between someone who chooses a variety compared to someone who closes themselves and who gets comfort in their relationships with things and people they know. To make it easier to get to know someone or to understand big five personalities can be seen in the following table where the difference in character between low scores and high scores.

#### D. Text mining

Text mining is a variation of data mining that seeks to find interesting patterns from a large collection of textual data. *Text mining* from other sources is a process of combining information extracted from various sources. Completion of *text mining*, there are several stages in general [9].

Text Preprocessing the initial stages of text mining to prepare text into data that will be processed at a later stage. In-text mining, raw data that contains information has an arbitrary structure, so the process of converting forms into structured data as needed is required, which is usually going to be numeric values. This process is called *Text Preprocessing*.

Tokenizing is the process of separating every word in a sentence into a single word. Each word in the sentence is separated using a space character. In this process, apart from separating words in sentences, a lowercase process is carried out. *Lowercase* is a process to change all letters in a sentence to lowercase, meaning that if there are capital letters in the sentence, this process will automatically change the capital letters into *lowercase*. In this process, only letters "a" to "z" are accepted, in addition to letters "a" to "z" such as characters, numbers, and punctuation (punctuation) are removed.

Feature Selection or feature selection, this stage aims to delete words that are considered not important. Feature selection has two stages, stopword and stemming.

1. Stopword

Stopwords are common words or words that often appear but do not have meaning, often used in sentences or documents that will cause a value to be small in helping to choose documents that suit your needs. In this process, what will be done is to erase common words to reduce the number of occurrences of words that have no meaning. Some words that often appear like "that", "and", "like", "often", "in", "will", "in", "this", "that", "no", "yes", "Is", "is", "ie", "ie", "in", "where", "for", "of", "and", "again", "you", "I", "then", "Then", "can", "to", "a", "like", "then", "said", "can", "so", "with", etc. which are considered to have no meaning means it.

2. Stemming

Stemming is a process of returning words in a sentence or document to their basic word form. The process of removing all affixes (*affix*) consists of a prefix (*prefix*), insert (*infix*), suffix (*suffix*) and duplication. Prefixes such as "on", "me", "air", "pen", "peng", "mem", and so forth. Inserts (*infix*) for example like "-er", "-el", "-em-", and "-in". While the examples of suffixes are "an", "right", "i", "it", "man", "wati", "wan", "asi", "ism", "in", "wi", and so forth.

E. Naïve Bayes Classifier

Naïve Bayes Classifier is an algorithm used to find the highest probability value for classifying test data in the most appropriate category put forward by Thomas Bayes as in equation (1) [9][10][12].

$$P(H|X) = \frac{p(X|H)p(H)}{P(X)} \tag{1}$$

Where:

X is the data with unknown classes

H is the data X hypothesis is a specific class.

P (H|X) is the probability of hypothesis H based on condition X (posterior probability).

P (H) is the probability of Hypothesis H (prior probability).

P (X / H) is the probability of X based on the conditions in Hypothesis H

P (X) is the probability of X

where the process Naïve Bayes Classifier method is as follows:

- Equation (2) is used in the Naïve Bayes Classifier Algorithm,

$$P(A | B) = (P(B | A) * P(A))/P(B) \tag{2}$$

- Event A on condition B is determined from opportunity B with condition A, opportunity A, and opportunity B. In the application later it becomes Equation (3),

$$P(Ci | D) = (P(D | Ci) * P(Ci))/P(D) \tag{3}$$

- Naïve Bayes Classifier, or biased referred to as multinomial Naïve Bayes, is a simplified model of the Bayes Theorem that fits in classifying Facebook Status categories. Therefore, eating equation (4),

$$Vmap = arg \max P(Vj|a1, a2, ... an) \tag{4}$$

- According to Equation (2), Equation (3) can be written as Equation (5),

$$Vmap = \frac{Arg \max_{Vj \in V} p(\alpha1, \alpha2 \dots an|Vj)P(Vj)}{P(\alpha1, \alpha2 \dots an)} \tag{5}$$

Because P (α1, α2, .....an is constant, so Equation (5) can be written as Equation (6),

$$Vmap = (Arg \max_{Vj \in V} p(\alpha1, \alpha2 \dots an | Vj)P(Vj)) \tag{6}$$

Where, Vj ∈ V, Vmap = The highest category probability, P(Vj) = Chance of the highest class type or j-category, and P(ai, a2, a3, ..... an|Vj) = Opportunity attribute if the known state of Vj

However, because P (a1, a2, a3 ..... ..an | vj) is difficult to calculate, it will be assumed that each word has no relation, indicated by Equation (7).

$$VMAP = Arg \max_{Vj \in V} \prod P (ai | Vj) \tag{7}$$

Vj ∈ V. So the Naive Bayes classifier calculation with Equation (8):

$$P(a1|Vj) = \frac{nci+m.p}{n+m} \tag{8}$$

Where:

nci is the i-th variable class word that has a yes or no value (1 or 0)

P is 1/ Number of Classes

$P$  is the opportunity

$m$  is the number of parameters / total words

$n$  is the number of word records in each class category

Equation (8) is resolved through the following calculation:

1. Specifies the value of  $n_c$  for each class
2. Calculate the value of  $P(a_i | V_j)$  and calculate the value of  $P(V_j)$  using Equation (8)  

$$VMAP = \text{Arg Max } P(V_j) \prod P(a_i | V_j) \quad (8)$$

$$V_j \in V$$
3. Calculates  $P(a_i | V_j) \times P(V_j)$  for each high class
4. Determine the classification results, namely the high class which has the greatest multiplication results

### III. RESULT AND DISCUSSION

Personality data in Fig.2(a) that is input will appear on the admin personality menu page can also edit or delete to the database. The personality data menu page is used to enter personality data to be used as a classification. On the personality data edit page, you can change only the type table in Fig.2(b). For personality ids, it cannot be changed because personality ids are auto-added.

Fig 2(a). Personality Data Edit Form

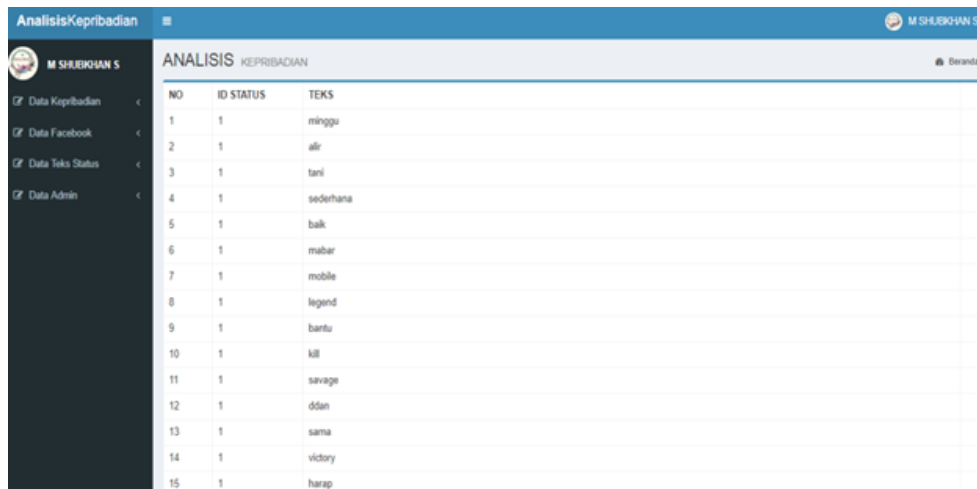
Fig 2(b). Personality Input Form

In this Facebook data page menu in Fig.3 (a), the admin can add Facebook data, by selecting the add button at the top. In the Facebook data, input form admin enters the Facebook name data by the user ID name on Facebook, to input admin status enter the pure status text does not need to be deleted or added, although the status of the word used is not by the spelling and symbols - symbols that are used do not need to be deleted. Because the status data entered will be processed with text mining and will produce basic words. Then the admin enters the personality that has been inputted on the personal input form, and the admin just has to choose the dropdown button that matches his personality as seen in Fig.3(b).

Fig 3(a). Facebook Data Menu Page

Fig.3 (b). Form Input Data Facebook

On the Facebook data menu page on the right, there is an edit and delete link. In the edit can not be used because the data input is valid and has obtained the personality category obtained from the questionnaire results and has been calculated by the psychologist method. To delete the link can be used because it will delete the whole meaning that the data is no longer entered into the database. The Status Data page in Fig.4 contains words from all statuses in the training data that has been through the text processing process, which in the process produces standard words or important words that are used to analyze a person's personality.



NO	ID STATUS	TEKS
1	1	minggu
2	1	air
3	1	tari
4	1	sederhana
5	1	baik
6	1	mabar
7	1	mobile
8	1	legend
9	1	bantu
10	1	kill
11	1	savage
12	1	ddan
13	1	sama
14	1	victory
15	1	harap

Fig.4. Data Status Menu Page

The evaluation result based on table I, it can be seen that the Naïve Bayes classifier has an accuracy value of 88%, and 22 users are true or by their personalities, and three other users are false or not by their personalities.

$$\text{accuracy} = (\text{the amount of data is correct}) / (\text{the total amount of data}) \times 100\% = 22/25 \times 100\% = 88\%.$$

Table I.  
 Trial Results

No	Facebook accounts	Personality	Personality System	B/S
1	Respondent 1	Extraversion	Extraversion	B
2	Respondent 2	Neuroticism	Neuroticism	B
3	Respondent 3	Agreeableness	Agreeableness	B
4	Respondent 4	Agreeableness	Agreeableness	B
5	Respondent 5	Openness	Openness	B
6	Respondent 6	Neuroticism	Neuroticism	B
7	Respondent 7	Neuroticism	Neuroticism	B
8	Respondent 8	Neuroticism	Neuroticism	B
9	Respondent 9	Conscientiousness	Conscientiousness	B
10	Respondent 10	Openness	Openness	B
11	Respondent 11	Openness	Openness	B
12	Respondent 12	Extraversion	Extraversion	B
13	Respondent 13	Agreeableness	Agreeableness	B
14	Respondent 14	Agreeableness	Agreeableness	B
15	Respondent 15	Extraversion	Conscientiousness	S
16	Respondent 16	Extraversion	Extraversion	B
17	Respondent 17	Openness	Openness	B
18	Respondent 18	Neuroticism	Neuroticism	B
19	Respondent 19	Conscientiousness	Openness	S
20	Respondent 20	Conscientiousness	Conscientiousness	B
21	Respondent 21	Openness	Openness	B
22	Respondent 22	Extraversion	Extraversion	B
23	Respondent 23	Agreeableness	Agreeableness	B
24	Respondent 24	Conscientiousness	Neuroticism	S
25	Respondent 25	Conscientiousness	Conscientiousness	B

#### IV. CONCLUSION

The BigFive psychologist theory that reflects the five personality categories can determine the personality category of each student by looking at the results of the questionnaire that is filled out. From these results, students are more precise on how to adapt to the environment around the university. Naïve Bayes classifier is more appropriate to use because the accuracy obtained from Facebook social media data, which is divided into two data between training data and testing data, produces a perfect accuracy value with a value of 88%. For further research development, it might be possible to use the platform assistance from the newer Facebook API to make it easier, obtaining data, and obtaining more data so that the accuracy value generated is higher. In future studies, it may be possible to use other social media by taking data uploaded in the form of images to find out the person's character/personality.

#### REFERENCES

- [1] M. Mulawarman and A. D. Nurfitri, "Perilaku Pengguna Media Sosial beserta Implikasinya Ditinjau dari Perspektif Psikologi Sosial Terapan," *Bul. Psikol.*, vol. 25, no. 1, pp. 36–44, 2017.
- [2] H. Sari, "HUBUNGAN TIPE KEPRIBADIAN DENGAN MOTIVASI BELAJAR PADA FAKULTAS KEPERAWATAN UNIVERSITAS SYIAH KUALA The Correlation Between Personality Type and Motivation to learn in Student with Competency-Based Curriculum," vol. VII, no. 2, pp. 1–10, 2016.
- [3] N. Susanti, "Aplikasi Psikologi Berbasis Web Untuk Menentukan Jenis Pekerjaan Sesuai Dengan Kepribadian Seseorang Menggunakan Teori MYERS-BRIGGS TYPE INDICATOR," *Maj. Ilm. Inform.*, vol. 3, no. 2, pp. 183–195, 2012.
- [4] I. Setiawati, G. Abdillah, and A. I. Hadiana, "Sistem Pendukung Keputusan Penempatan Karyawan Berdasarkan Test Dominant-Influence-Steady-Compliance (Disc) Menggunakan Metode Profile Matching," *Pros. SNST Fak. Tek.*, vol. 1, no. 1, pp. 25–30, 2016.
- [5] R. A. N. Prasasti, "Hubungan Antara Dimensi Kepribadian Big Five Dengan Perilaku Merokok Pada Remaja Akhir," 2011.
- [6] M. Z. Sarwani, D. A. Sani, and F. C. Fakhri, "Personality Classification through Social Media Using Probabilistic Neural Network Algorithms," *Int. J. Artif. Intell. Robot.*, vol. 1, no. 1, p. 9, 2019.
- [7] E. N. Nugrahawati and D. Arista, "Profil Kepribadian Big Five dan Collecting Behaviour Kolektor Piringan Hitam," *Psymphatic J. Ilm. Psikol.*, vol. 3, no. 1, pp. 69–76, 2017.
- [8] M. Z. Sarwani and W. F. Mahmudy, "Analisis Twitter Untuk Mengetahui Karakter," *Semin. Nas. Sist. Inf. Indones.*, no. November, pp. 2–3, 2015.
- [9] J. N. Apriliana, Natalis Ransi, "Implementasi Text Mining Klasifikasi Skripsi Menggunakan Metode Naïve Bayes Classifier," *Semant. Vol.3, No.2, Jul-Des 2017*, vol. 3, no. 2, pp. 329–343, 2017.
- [10] G. A. Buntoro, "Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naive Bayes Classifier Dan Support Vector Machine," *Din. Inform.*, vol. 5, no. September 2016.
- [11] J. I. Syariah, "Pengaruh Media Sosial Terhadap Kepribadian Mahasiswa STEBIS IGM Palembang \*," vol. 4, no. 2, pp. 327–336, 2016.
- [12] Alfa Saleh, "Implementasi Metode Klasifikasi Naïve Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec J.*, vol. 2, no. 3, pp. 207–217, 2015.