

Personality Prediction Based on Facebook Media Social Status Using the Method Naïve Bayes and KNN



Andi Oktafiqurahman¹, Kusri², Asro Nasiri³

^{1,2,3}MTI Universitas AMIKOM Yogyakarta, Jl Ringroad Utara, Condongcatur, Depok, Sleman, Yogyakarta Indonesia 55283

ABSTRACT: Social media is now almost a part of life for people of almost all ages with diverse personalities. This study aims to predict a person's personality based on Facebook social media status using the Naïve Bayes algorithm and K-NN personalities based on the big five. The data taken from *MyPersonality* which is divided into 40% training data, and 60% test data produces an accuracy rate of 100%, 100% precision, 100% recall can be concluded that the prediction variables are (significant) with the Naïve Bayes method. While predictions using the KNN method showed an average personality prediction value with an accuracy of 58.96% (Insignificant), precision 99.12% (Significant), and recall 2.34% (Insignificant). Based on research can be concluded that both Naïve Bayes and KNN methods can be used to predict personality based on Facebook status.

KEYWORDS: Naïve Bayes, K-NN, Prediction, Personality, Facebook

INTRODUCTION

Social media in Indonesia over time has grown rapidly. This is supported by increasingly advanced internet infrastructure, such as wifi access, fiber networks, and so on (Istiani & Islamy, 2020). Indonesian people are increasingly active in the world of social media, with a penetration rate that reaches tens of millions of users, almost all groups have social media both men and women. This media was created to facilitate the community in carrying out difficult activities carried out by humans (Astuti et al., 2020). The existence of social media certainly has an impact, both positive and negative so that any content can be viral easily (Herlina, 2018).

The advent of the internet has changed everything from the way of life to the way of thinking (Simangunsong, 2016). Now everyone can publish their thoughts in the public domain by simply posting a status on social media. On the one hand, the existence of this media can help for positive things of human progress. But on the other hand, it can have a negative impact, namely about unlimited freedom. Everyone can freely publish whatever is on their minds, be it opinions, criticisms even to blasphemy and insults. The things or themes published are also very diverse ranging from daily activities, politics, culture, and even religion (Tuty Mutiah, Ilham Albar, Fitriyanto, 2019).

Social media creates interactions with fellow users just like that with people we know or with new people who didn't know each other before (Yasya et al., 2019).. Social media is a new form of information technology that is growing rapidly and its users are always increasing rapidly every day. With the rapid development of technology now social media always has a changing form and type but nowadays social media has several forms and types such as weblogs ("blogs"), microblogs, social networking sites, and media-sharing, RSS and other web syndication feeds, wikis, social bookmarks, mashups, widgets, microblogging, and so on (Sherlyanita & Rakhmawati, 2016). The internet users worldwide have reached 31.7 billion and from year to year the number of internet users has grown to 7.6 percent. As for social media users themselves reached 2.2 billion with users reaching 3.7 billion (Sherlyanita & Rakhmawati, 2016).

In the world of psychology there are several personalities owned by humans themselves, including *Sanguine*, *choleric*, *Melancholy*, *plegmatic*. Where are people of type *Sanguine* It has a more open and sociable nature, while *choleric* has a more logical nature, and *Melancholy* tends to have a perfectionist personality, sensitive, empathetic, and tends to be reserved, to *plegmatic* himself has a calmer nature and has high empathy (Seto Mulyadi, Warda Lisa, 2019).

There are once words that can describe a person's personality. Almost every one of us uses when interacting or communicating with others as social beings. We often say that A is like this and B is like that or maybe A has "some personality" and B has no personality at all"(Wardani, 2019). We use the word personality as we please. Many times, we are actually just describing a certain disposition or personality. The idea that we have several different personalities is not quite right. However, it is very interesting. In communicating on social media, we don't have to see people directly, we often set someone's personality the same or different from ourselves, where basically or better we say, everyone has their own personality patterns(Seto Mulyadi, Warda Lisa, 2019).

METHODS

This study aims to predict what personality users have from Facebook social media using the naïve bayes and k-nearest neighbor methods. By making predictions based on Facebook social media status. By knowing someone's personality, we will be younger to interact or also make it easier for companies that want to know the personality of prospective employees from social media whether their personality runs parallel to the company's vision and mission. As we know in the world of work, a person's personality can have an impact on the results of his work. Naïve bayes and k-nearest neighbor algorithms can be used to make such predictions. The naïve bayes algorithm is very popular in research conducted to make predictions. And the advantage of this algorithm is that it does not require so much training data to determine the estimated parameters used in the classification process. Likewise, the K-nearest neighbor algorithm is quite popular because this algorithm can be used in the classification and regression process. Can be used in a diverse number of classes. It is very easy if there is additional data because new data can be directly classified and does not require a training process and model building process, and so on.

The objectives in writing this research are as follows: 1) Knowing a person's personality from facebook status using naïve bayes and k-nearest neighbor algorithms, 2) as a requirement to obtain a master's degree in computer at the faculty of computer science, informatics at Amikom University Yogyakarta. 3) As a development of knowledge obtained at Amikom University Yogyakarta

In writing this research, the limitations of the problem will be formulated as follows: 1) The research was conducted based on the status of Facebook social media users. 2) Create a personality prediction system based on social media status using the big five personality traits approach model. 3) Using naïve bayes and k-nearest neighbor algorithms to create a prediction system.

RESULT

1. Big Five Personality

Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness are traits of high-level personality that encompass The Big Five theory. The traits of the big five personalities are orthogonal and bipolar, which from individual to individual profile in the big five can be unique. With the exception of openness to experiences that arise later in adolescence, some researchers further argue that the big five are innate traits of genetic disposition that do not change over time. However, there is most likely a transactional relationship between personality and social contextual factors. The personality traits of the big five are divided into 5 with which:

a. Extraversion

This personality dimension relates to a person's comfort level in interacting with others. The positive characteristics of extraversion individuals are sociable, sociable, group life and assertive

b. Agreeableness

These individuals tend to be obedient to other individuals and have personalities that want to avoid confrontation. The positive characteristics of this individual dimension are cooperative, trusting, kind, warm and gentle and helpful.

c. Conscientiousness

Individuals in this dimension have a more cautious nature in carrying out an action or very considerate in making a decision, this personality dimension also has a high self-discipline and trustworthy nature. The positive characteristics of this dimension are reliable, responsible, diligent, and achievement-oriented.

d. Neuroticism

This personality dimension has a person's ability to withstand pressure or stress. The positive characteristic of this dimension is called emotional stability. These emotionally stable individuals tend to be calm when facing problems, confident, have a firm stand.

e. Openness (Open)

This personality dimension groups individuals based on an interest in new things and a desire to know and learn something new. The positive characteristics of this personality dimension are more creative, imaginative, intellectual, curious and broad-minded.

2. Natural Language Processing (NLP)

Natural Language Processing is one part of Artificial Intelligence (AI) that allows computers to interact such as reading, understanding, and interpreting the intentions of humans using natural language. NLP helps measure sentiment and determine which parts are important parts of human language. For computers this is a difficult thing to do because of the large amount of data that is not structured, has no formal rules, and there is no correlation in the real world. Some of the common tasks that NLP performs are information extraction, document categorization, semantic text matching (Wijeratne et al., 2019).

3. Personality

In the world of psychology there are many personalities possessed by humans. Many words can describe a person's personality or disposition. Every day we almost use when interacting with others. We sometimes say that A is like this and B is like this or that A has "two personalities" or B has "no personality at all" or even has multiple personalities. Many times we use the word personality casually. However, we are actually only describing a certain disposition or number of personalities.

4. Naive Bayes

Personality Prediction Based on Facebook Media Social Status Using the Method Naïve Bayes and KNN

Naïve bayes is an algorithm commonly used to perform binary and multiclass classifications. This algorithm can assess models very quickly and scale linearly in the number of predictions and rows. This algorithm can also perform a classification that presents each class of objects based on conclusions or probabilistic recapitulation and find the most likely class that corresponds to each object to be determined class from existing test data based on attributes or variables that have known values. For the equation of naïve bayes see below:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

5. K-Nearest Neighbor (KNN)

KNN algorithm or K-nearest Neighbor is an algorithm used to aggregate data based on training data (training data) derived from k nearest neighbor (nearest neighbor). Where k is the number of nearest neighbors. The concept of how this algorithm works is to get the shortest distance between the data to be evaluated and k neighbors in the training data. The purpose of the K-Nearest Neighbor algorithm is to classify new objects based on attributes and sample training data. For the process of calculating distances is carried out using the Euclidean concept.

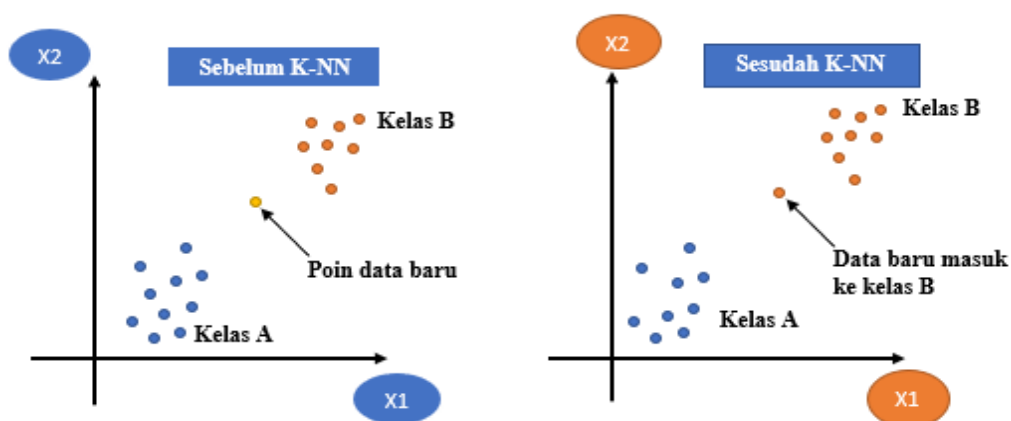


Figure 1. K-NN

6. Preprocessing Data

The dataset used in the study Personality Prediction Based on Facebook Social Media Status Using the Naïve Bayes Dan Knn Method consists of 9,990 Facebook Status Data from previous research from Kaggle Machine Learning which has been tested using Big Five Personality Trait with five variable criteria namely sEXT, sNEU, sAGR, sCON, sOPN.

7. Implementation of the Naïve Bayes Method

The next step after the Dataset is ready is then to apply the Naïve Bayes method using RapidMiner Studio Software. Naïve Bayes method was implemented to predict personality based on Facebook social media status. The Naïve Bayes model has been trained using training data that has gone through the preprocessing stage, in this study predicting variable criteria namely sEXT, sNEU, sAGRsCON, sOPN with the results of naïve bayes implementation can be seen in the following table:

Table 1. Results of the Naïve Bayes method

Variable	Accuracy	Precision	Recall
sEXT	100%	100%	100%
sNEU	99,95%	99,92%	100%
sAGR	99,89%	99,80%	100%
sCON	99,84%	99,89%	99,77%
sOPN	99,95%	99,80%	100%

8. KNN Method Implementation

The implementation of the KNN method was carried out to predict personality based on Facebook social media status. The KNN model has been trained using training data that has gone through the preprocessing stage, in this study predicting variable criteria, namely sEXT, sNEU, sAGR, sCON, sOPN with the results of KNN implementation can be seen in the following table;

Table 2. results of KNN implementation

Variable	Accuracy	Precision	Recall
sEXT	57,11%	100%	0,49%

Personality Prediction Based on Facebook Media Social Status Using the Method Naïve Bayes and KNN

sNEU	38.46%	100%	2.18%
sAGR	47,04%	100%	0,40%
sCON	57,16%	95,59%	7,45%
sOPN	74,54%	100%	1,03%

9. Comparison of Results Between Naïve Bayes Method and K-NN

Naive Bayes: The Naive Bayes method is based on Bayes' Theorem, which uses conditional probability to perform classifications. This method assumes that all features or variables are conditionally independent. K-Nearest Neighbors (KNN): The KNN method is based on the principle that similar objects tend to be in the same category. This algorithm looks for K closest neighbors of the object to be classified and uses the majority of those neighbor categories to classify that object. Based on the implementation of the Naïve Bayes Method and KNN that has been carried out, there is a significant comparison of results, which can be seen in the following diagram;

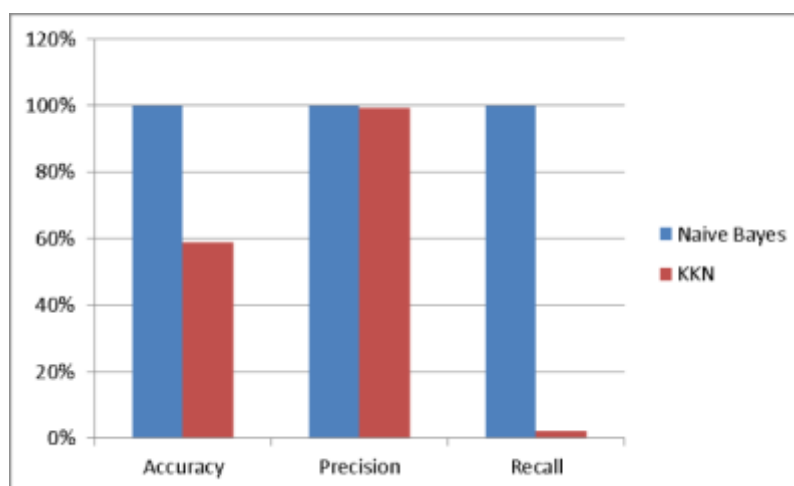


Figure 2. Comparison of Results

The implementation of the Naive Bayes method as a personality prediction method based on Facebook social media status obtained a significant average value with an accuracy value of 100%, Precision 100%, and a Recall value of 100%. This is inversely proportional to the results of the implementation of the KNN method where the average accuracy result was 58.69%, precision 99.12% and recall value 2.34.

Table 3. the results of the implementation of the KNN method

Method	Accuracy	Precision	Recall
Naive Bayes	100%	100%	100%
KKN	58,96%	99,12%	2,34%

DISCUSSION

Research conducted by Mohammad Ichsanudin A (2021) entitled Personality prediction based on twitter social media using the method Naïve Bayes classifier With the aim of producing personality predictions with Twitter social media status with the prediction model used, namely: Big Five. Resulting in the conclusion of this study that the resulting accuracy is quite low with an accuracy of only 42%. What is obtained from the prediction results label The highest personality is found in Label Agreeableness with the number of 167. Meanwhile, the lowest value is obtained by Extraversion label with a total of 14 tweets. However, in the study, there is a weakness, namely with the number of Dataset Not much is used because not all accounts do not have tweets or tweets of more than 100 tweets (Ichsanudin et al., 2021).

Research conducted by Sandhya katiyar (2020) entitled personality prediction from stack overflow by using naïve bayes theorem in data mining. This study aims to analyze a person's personality on stack overflow forums to build a team in a project using the naïve bayes algorithm. The results of this study resulted in conclusions from 16 personality types based on the MBTI obtained the results of the ENTP which defines characteristics such as honest, assertive and having leadership qualities. The study has similarities with the authors, namely making personality predictions and using naïve bayes algorithms, but the social media used is different and the number of personalities used is different (Katiyar et al., 2020)

Personality Prediction Based on Facebook Media Social Status Using the Method Naïve Bayes and KNN

Research conducted by Siska Febriani (2021) entitled Analysis of diagnostic data for classification of personality disorders using the C4.5 algorithm. With the aim of predicting personality disorders to get the information needed by society. In this study only used 3 psychological disorders, namely anxiety disorders (GAD), mood disorders (major depressive disorder), and somatoform disorders (conversion disorders) from this study resulted in an accuracy rate of 72.67%. The similarity between this study and the author is to do personality prediction, it's just that the study did predict personality disorders and also used the C4.5 algorithm, while the author used naïve bayes and k-nearest neighbor (Febriani & Sulistiani, 2021)

Research conducted by Irina Deeva (2019) entitled "Computational Personality Prediction Based on Digital Footprint of A Social Media User" in this study determines whether psychometrics can correlate equally well in different media environments with this goal a correlation matrix is built between features and psychometrics. With the data divided into 30% test data and 70% training data to avoid repeated training by the model. With it train-test-split method with hyperparameter shuffle = 42. Univariate regression results are presented for facebook. As for the research that will be carried out is to make predictions where comparing the data that has been predicted with a deep learning approach model for modeling the big five personality traits personality with a machine learning approach model (Deeva, 2019)

Research conducted by Hans Christian (2021) entitled "Text based personality prediction from multiple social media data sources using pretrained language model and model averaging" in this study shows a comparison of different feature extraction methods with different algorithmic approaches in building personality systems for Facebook social media data sources. This research conducts an architectural deep learning approach with BERT, RoBERTa, and XLNet language models. The statistical features and model averages outperformed most personality models by producing the highest accuracy of 86.17% and an f1 measurement score of 0.912 on the Facebook dataset and an accuracy of 88.49% and an f1 measurement score of 0.882 on the Twitter dataset. Meanwhile, the research will be carried out using machine learning methods using the naïve bayes algorithm and k-nearest neighbor (Hans et al., 2021).

CONCLUSION

Based on research can be concluded that both Naïve Bayes and KNN methods can be used to predict personality based on Facebook status. Personality prediction using the Naive Bayes method shows the average value of personality prediction with 100% accuracy, 100% precision, 100% recall can be concluded that the prediction variables are (Significant) with the Naive Bayes method. While predictions using the KKN method showed an average personality prediction value with an accuracy of 58.96% (Insignificant), precision 99.12% (Significant), and recall 2.34% (Insignificant).

REFERENCES

- 1) Astuti, M., MM, M. O. S., Matondang, N., Kom, S., & MM, M. T. (2020). *Manajemen Pemasaran: UMKM dan Digital Sosial Media*. Deepublish.
- 2) Deeva, I. (2019). Computational Personality Prediction Based on Digital Footprint of A Social Media User. *Procedia Computer Science*, 156, 185–193. <https://doi.org/10.1016/j.procs.2019.08.194>
- 3) Febriani, S., & Sulistiani, H. (2021). Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5. *89Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 2(4), 89–95.
- 4) Hans, C., Suhartono, D., Andry, C., & Zamli, K. Z. (2021). Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. *Journal of Big Data*, 8(68).
- 5) Herlina, L. (2018). Disintegrasi Sosial Dalam Konten Media Sosial Facebook. *TEMALI: Jurnal Pembangunan Sosial*, 1(2), 232–258. <https://doi.org/10.15575/jt.v1i2.3046>
- 6) Ichsanudin, M., Susilo, A., Irawan, Y., Solehudin, A., Studi, P., Informatika, T., Karawang, S., Ronggo, J. H., Teluk, W., & Karawang, J. T. (2021). Prediksi Kepribadian Berdasarkan Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 988–996.
- 7) Irawan, A. S. Y. (2021). Prediksi Kepribadian Berdasarkan Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 5(2), 988-996.
- 8) Istiani, N., & Islamy, A. (2020). Fikih Media Sosial Di Indonesia. *Asy Syar'iyah: Jurnal Ilmu Syari'Ah Dan Perbankan Islam*, 5(2), 202–225. <https://doi.org/10.32923/asy.v5i2.1586>
- 9) Katiyar, S., Kumar, S., & Walia, H. (2020). Personality prediction from stack overflow by using Naïve Bayes theorem in data mining. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9(3).
- 10) Seto Mulyadi, Warda Lisa, A. N. K. (2019). Psikologi kepribadian kognitif. In *Psikodimensia* (Vol. 18, Issue 1).
- 11) Sherlyanita, A. K., & Rakhmawati, N. A. (2016). Pengaruh dan Pola Aktivitas Penggunaan Internet serta Media Sosial pada Siswa SMPN 52 Surabaya. *Journal of Information Systems Engineering and Business Intelligence*, 2(1), 17. <https://doi.org/10.20473/jisebi.2.1.17-22>
- 12) Simangunsong, B. A. (2016). Interaksi Antarmanusia melalui Media Sosial Facebook Mengenai Topik Keagamaan. *Jurnal ASPIKOM*, 3(1), 65. <https://doi.org/10.24329/aspikom.v3i1.99>

Personality Prediction Based on Facebook Media Social Status Using the Method Naïve Bayes and KNN

- 13) Tuty Mutiah, Ilham Albar, Fitriyanto, A. R. (2019). Etika Komunikasi Dalam Menggunakan Media Sosial. *Global Komunika*, 1(1), 14–24.
- 14) Wardani, L. (2019). *Gambaran Kesejahteraan Psikologi Generasi Milenial Pengguna Media Digital Di Jakarta Laila Meiliyandrie Indah Wardani & * Sri Wahyuning Astuti Faculty of Psychology , Mercu Buana Universiti * Correspondent author ' s e -mail : Generasi milenium mewakili h. 7(1), 1–14.*
- 15) Wijeratne, Y., Silva, N. de, & Shanmugarajah, Y. (2019). Natural Language Processing for Government: Problems and Potential. *Building Chatbots with Python, August*, 29–61.
- 16) Yasya, W., Muljono, P., Seminar, K. B., & Hardinsyah, H. (2019). Pengaruh Penggunaan Media Sosial Facebook Dan Dukungan Sosial Online Terhadap Perilaku Pemberian Air Susu Ibu. *Jurnal Studi Komunikasi Dan Media*, 23(1), 71. <https://doi.org/10.31445/jskm.2019.1942>



There is an Open Access article, distributed under the term of the Creative Commons Attribution – Non Commercial 4.0 International (CC BY-NC 4.0) (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits remixing, adapting and building upon the work for non-commercial use, provided the original work is properly cited.