

## Klasifikasi Kepribadian Pengguna Media Sosial

Enie Yuliani<sup>1</sup>, Ema Utami<sup>2</sup>, Suwanto Raharjo<sup>3</sup>

<sup>12</sup>Magister Teknik Informatika Universitas Amikom Yogyakarta

<sup>3</sup>Institut Sains & Teknologi AKPRIND

Email: <sup>1</sup>enie.yuliani@students.amikom.ac.id, <sup>2</sup>ema.u@amikom.ac.id, <sup>3</sup>wa2n@nrar.net

### Abstrak

Media sosial merupakan tempat para penggunanya berbagi hal tentang diri mereka dan pandangan hidup penggunanya. Beberapa penelitian yang memanfaatkan media sosial sebagai objeknya telah dilakukan karena menyadari potensi banyaknya informasi yang bisa digali dari media sosial. Tujuan dari penelitian ini adalah untuk mengetahui apakah mungkin melakukan prediksi kepribadian berdasarkan media sosial dan melakukan pemodelan klasifikasi kepribadian berdasarkan media sosial. Proses dalam melakukan text mining mulai dari pengumpulan data, pembagian dataset, pre-processing teks, ekstraksi fitur hingga pemilihan algoritma klasifikasi berpengaruh terhadap akurasi prediksi kepribadian yang dilakukan. Memadukan fitur perilaku sosial pengguna media sosial dan fitur linguistik psikologis dapat meningkatkan akurasi dari klasifikasi.

**Kata kunci:** klasifikasi kepribadian, text mining, prediksi kepribadian media sosial

## 1. PENDAHULUAN

Mengetahui kepribadian seseorang merupakan sesuatu yang penting karena kepribadian dapat mempengaruhi tingkah laku seseorang yang dapat berpengaruh pula terhadap tindakan pengambilan keputusannya. Oleh karena itu kepribadian seseorang menjadi salah satu bahan pertimbangan bagi HRD pada saat proses seleksi calon karyawan. Penelitian dibidang psikologi menemukan bahwa ada hubungan antara kepribadian seseorang dengan cara mereka menulis atau berbicara. Orang dengan kepribadian yang sama cenderung menggunakan kata yang sama untuk mengekspresikan diri (Lin & Mao, 2015). Pemilihan kata dapat mencerminkan kepribadian seseorang (Mairesse, Walker, Mehl, & Moore, 2007).

Media sosial merupakan tempat para penggunanya berbagi hal tentang diri mereka dan pandangan hidup penggunanya. Beberapa penelitian yang memanfaatkan media sosial sebagai objeknya telah dilakukan karena menyadari potensi banyaknya informasi yang bisa digali dari media sosial. Termasuk penelitian tentang kepribadian.

Berbagai metode digunakan guna mendapatkan hasil akurasi dan model terbaik

untuk melakukan klasifikasi kepribadian berdasarkan media sosial.

Berdasarkan hasil penelitian-penelitian sebelumnya mengenai klasifikasi kepribadian dengan memanfaatkan media sosial, diperoleh hasil bahwa media sosial dapat menjadi alat identifikasi kepribadian seseorang. Namun untuk memperoleh hasil akurasi terbaik perlu dilakukan studi lebih lanjut. Penelitian ini akan mereview metode dan langkah dari beberapa penelitian mengenai klasifikasi kepribadian pengguna media sosial sehingga dapat dimanfaatkan untuk penelitian selanjutnya.

## 2. METODE PENELITIAN

### a. Gambaran Umum

Secara umum, tahapan-tahapan yang dilakukan antara lain:

- 1) Pengumpulan dataset
- 2) *Pre-processing Text*
- 3) Ekstraksi Fitur
- 4) Model Klasifikasi

### b. Data

Berdasarkan data dari *pers release* platform media sosial yang dirangkum oleh hootsuite.com, total pengguna media sosial di

Indonesia tahun 2019 adalah sebanyak 150 juta pengguna. Jumlah tersebut mencapai 56% dari jumlah total penduduk di Indonesia. Sejalan dengan populernya media sosial tersebut, penelitian mengenai kepribadian berdasarkan media sosial tersebut juga telah banyak dilakukan. (Rumagit & Girsang, 2018) memprediksi kepribadian dari pengguna facebook dengan memanfaatkan API facebook. Penelitian tersebut melibatkan 345 partisipan. Klasifikasi kepribadian dengan menggunakan platform twitter juga berhasil dilakukan oleh (Lukito, Erwin, Purnama, & Danoekoesoemo, 2016) dengan total dataset sebanyak 240.000 tweet dengan rata-rata 2.500 tweet per pengguna.

Dalam proses text mining, dataset yang ada perlu dibagi menjadi data latih dan data uji. Dibandingkan dengan porsi 50% banding 50% dan 60% banding 40%, porsi pembagian data latih dan data uji sebanyak 70% banding 30% lebih direkomendasikan karena memiliki akurasi paling tinggi yaitu sebesar 84,00% (Lhaksmana, Nhita, & Anggraini, 2017).

### c. Pre-processing Teks

Data yang belum terstruktur perlu diubah menjadi data yang terstruktur sesuai dengan kebutuhan sebelum proses penambahan data yang lebih, proses ini disebut *pre-processing* teks (Naradhipa & Purwarianti, 2012). Tahapan yang umum dilakukan pada pre processing teks ini antara lain :

- 1) *Case folding*  
Tahapan ini menangani adanya huruf kapital dan huruf kecil. Pada tahap ini semua huruf diubah menjadi huruf kecil (*lowercase*).
- 2) *Tokenizing*  
Dalam proses ini, kalimat atau sekumpulan kata akan dipecah menjadi kata atau frasa.
- 3) *Filtering*  
Kata yang didapat dari hasil *crawling* perlu difilter sehingga yang tersisa adalah kata yang penting saja. Algoritma untuk filtering dapat menggunakan algoritma *stoplist* (membuang kata kurang penting) atau algoritma *wordlist* (menyimpan kata penting).
- 4) *Stemming*  
Proses *stemming* merupakan proses mencari kata dasar dari suatu kata atau frasa. Untuk kasus teks mining dalam Bahasa Indonesia,

proses *stemming* sangat penting dilakukan karena Bahasa Indonesia memiliki *prefixes*, *suffixes*, *infixes* dan *confixes* yang membuat kata dasar dapat berubah menjadi banyak bentuk dan proses pencarian kata dasar menjadi cukup sulit (Asian, Williams, & Tahaghoghi, 2005).

### 5) Text normalization

Data yang diperoleh dari hasil *crawling* terutama jika data diperoleh dari media sosial, tidak selalu bentuknya merupakan bentuk yang normal. Kemungkinan akan terdapat kata tidak lazim seperti singkatan atau kata tidak baku. Maka diperlukan tahap *normalization*.

### d. Ekstraksi Fitur

Penelitian sebelumnya mengenai prediksi kepribadian pengguna media sosial menggunakan beragam fitur. Perilaku sosial dan fitur linguistik yang diperoleh dari status pengguna di media sosial dapat dijadikan fitur untuk memprediksi kepribadian. Untuk mendapatkan fitur apa saja yang memiliki pengaruh signifikan terhadap prediksi kepribadian dari media sosial, perlu dilakukan pengujian.

### e. Klasifikasi Kepribadian

Berbagai macam algoritma telah diujicobakan dalam penelitian-penelitian sebelumnya untuk mengetahui algoritma mana yang lebih tepat dan memiliki tingkat akurasi paling tinggi untuk diterapkan dalam kasus klasifikasi kepribadian berdasarkan media sosial.

## 3. TINJAUAN PUSTAKA

### a. Kepribadian Big Five

Kepribadian big five merupakan salah satu metode yang terkenal dalam dunia psikologi untuk menginterpretasikan kepribadian seseorang. Kepribadian big five terdiri dari (Golbeck, 2011):

- 1) Openness to Experience: sifat ini menilai keinginan individu untuk mencari dan menghargai pengalaman baru serta kecenderungan untuk senang mengetahui sesuatu yang tidak familiar (Pervin & John, 2001).
- 2) Conscientiousness: sifat ini menilai kemampuan individu dalam organisasi, baik mengenai tingkat keteraturan maupun motivasi dalam mencapai tujuan (Pervin & John, 2001).

- 3) Extraversion: sifat ini menilai intensitas interaksi interpersonal, tingkat ketergantungan dengan orang lain dan kemampuan untuk berbahagia.
- 4) Agreeableness: Orang yang memiliki skor tinggi pada sifat ini adalah orang yang lembut hati, dapat dipercaya, suka menolong, pemaaf, penurut.
- 5) Neuroticism: Sifat ini menggambarkan stabilitas emosional dengan cakupan perasaan negatif yang kuat termasuk kecemasan, kesedihan, iritabilitas dan ketidakpercayaan diri.

#### b. Text Mining

Text mining secara umum dapat didefinisikan sebagai proses menggali informasi dimana pengguna berinteraksi dengan sekumpulan dokumen menggunakan *tools* analisis yang merupakan komponen-komponen dalam data mining (Feldman & Sanger, 2009).

Proses pokok dalam text mining antara lain pemrosesan awal (*pre-processing*), transformasi teks (*text transformation*), pemilihan fitur (*feature selection*), dan penemuan pola text mining atau data mining (*pattern discovery*). (Even-Zohar, 2002).

### 4. HASIL DAN PEMBAHASAN

#### a. Algoritma Klasifikasi

Pemilihan algoritma klasifikasi akan berpengaruh terhadap hasil akurasi dan tingkat keberhasilan prediksi kepribadian seseorang berdasarkan media sosialnya. Berikut ini penulis rangkum perbandingan beberapa algoritma yang digunakan untuk melakukan klasifikasi kepribadian. (Lhaksana et al., 2017) melakukan klasifikasi kepribadian pengguna facebook menggunakan algoritma Backpropagation Neural Network dan memperoleh akurasi sebesar 84,00%.

(Rumagit & Girsang, 2018) mengklasifikasi kepribadian pengguna facebook dengan algoritma SVM, Naive Bayes dan Logistic Regression memperoleh hasil akurasi rata-rata sebesar 89,00%.

Algoritma KNN digunakan oleh (Ellandi, Setiawan, & Nugraha, 2019) untuk mengklasifikasi kepribadian pengguna twitter ke

dalam kelompok kepribadian Big Five dan memperoleh akurasi sebesar 60,97%.

Penelitian mengenai klasifikasi kepribadian pengguna facebook yang dilakukan (Tandera, Suhartono, Wongso, & Prasetyo, 2017) membandingkan kinerja *traditional machine learning* dan *deep learning*. Untuk *traditional machine learning* penulis menggunakan algoritma naive bayes, SVM, logistic regression, gradient boosting dan LDA. Untuk *deep learning* penulis menggunakan algoritma MLP, LSTM, GRU, CNN 1D, LSTM+CNN 1D). Akurasi tertinggi untuk *traditional machine learning* adalah dengan algoritma SVM sebesar 67,20%. Akurasi tertinggi untuk *deep learning* adalah dengan algoritma LSTM+CNN 1D sebesar 74,17%.

#### b. Ekstraksi Fitur

Penelitian yang dilakukan oleh (Ellandi et al., 2019) menguji kombinasi fitur yang memiliki pengaruh terhadap prediksi kepribadian pengguna media sosial. Menggunakan 15 fitur dari perilaku sosial para pengguna twitter yang antara lain jumlah followers, following, jumlah mention, jumlah hashtag, jumlah replay, jumlah URL, jumlah kata, jumlah retweet, jumlah media URL, jumlah tanda baca, jumlah emoji, rata-rata kata, jumlah huruf besar, jumlah karakter, rata-rata karakter.

Pada media sosial facebook, penelitian (Souri, Hosseinpour, & Rahmani, 2018) memanfaatkan perilaku sosial pengguna facebook sebagai variabel untuk klasifikasi kepribadian. Variabel yang digunakan antara lain jumlah *like*, *language*, *book*, *work*, *education*, *sport*, *activity*, *game*, *group*, *movie*, *music*, *friends*, *interested*, *links*, *TV shows*, *note*, *question*, *post*, *number of photo in the timeline*, *number of photo without text*, *newsfeed*, *show in timeline*.

Beberapa penelitian melakukan analisis fitur linguistik dengan menggunakan TFIDF dengan pendekatan *bag of words*, beberapa memanfaatkan LIWC, MRC, SPLICE, SNA untuk mengekstraksi fitur linguistik dari status pengguna di media sosial. Program-program tersebut merupakan program analisis teks yang dapat menghitung atau kata-kata dalam kategori psikologis.

Memanfaatkan LIWC untuk ekstraksi fitur dalam proses klasifikasi kepribadian pengguna facebook, (Lhaksana et al., 2017) menggunakan

*social word*, *positive emotions* dan *negative emotions* sebagai fitur.

Pemilihan fitur sangat berpengaruh terhadap hasil pemodelan dan hasil klasifikasi kepribadian. (Tandera et al., 2017) menggunakan 85 fitur dari LIWC dan 74 fitur dari SPLICE dalam melakukan klasifikasi kepribadian pengguna media sosial facebook.

### c. Analisis Hasil Perbandingan

Pada penelitian (Ellandi et al., 2019) hasil akurasi perpaduan fitur perilaku sosial dan fitur linguistik tidak sesuai target yang diharapkan karena adanya ketimpangan dataset. Pada 61 data latih dan 24 data uji, terdapat satu label kelas yang dominan sehingga model prediksi yang dibangun cenderung memprediksikan setiap keputusan adalah kelas yang dominan tersebut.

Penggunaan metode *deep learning* dalam klasifikasi kepribadian dalam beberapa studi terbukti memperoleh hasil akurasi yang lebih tinggi dibandingkan dengan metode *traditional machine learning*.

## 5. PENUTUP

Berdasarkan penelitian-penelitian mengenai klasifikasi kepribadian berdasarkan media sosial yang telah dilakukan sebelumnya, menilai kepribadian seseorang berdasarkan media sosialnya adalah sebuah hal yang mungkin untuk dilakukan. Proses dalam melakukan *text mining* mulai dari pengumpulan data, pembagian dataset, pre-processing teks, ekstraksi fitur hingga pemilihan algoritma klasifikasi berpengaruh terhadap akurasi prediksi kepribadian yang dilakukan.

Untuk membangun suatu model klasifikasi, keseimbangan jumlah dataset untuk masing-masing kelas perlu diperhatikan karena akan mempengaruhi hasil akurasi dalam klasifikasi.

## 6. REFERENSI

Asian, J., Williams, H. E., & Tahaghoghi, S. M. M. (2005). *Stemming Indonesian. Australian Computer Society Inc.*

Ellandi, R., Setiawan, E. B., & Nugraha, F. N. (2019). Prediksi kepribadian Big Five dengan Term-Frequency Inverse Document Frequency Menggunakan Metode k-Nearest

Neighbor pada Twitter. *E-Proceeding of Engineering*, 6.

Even-Zohar, Y. (2002). *Introduction to Text Mining*. Automated Learning Group National Center for Supercomputing Applications University of Illinois.

Feldman, R., & Sanger, J. (2009). *The Text Mining Handbook*. <https://doi.org/https://doi.org/10.1017/CBO9780511546914>

Golbeck, J. (2011). Predictiong Personality with Social Media. *Proc. 2011 Annu. Conf. Ext. Abstr. Hum. Factors Comput. Syst. CHI EA*, 253–262.

Lhaksmana, K. M., Nhita, F., & Anggraini, D. (2017). Klasifikasi Kepribadian Berdasarkan Status Facebook Menggunakan Backpropagation. *E-Proceeding of Engineering*, 4, 5174.

Lin, J., & Mao, W. (2015). Personality based public sentiment classification in microblog. *IEEE*.

Lukito, L. C., Erwin, A., Purnama, J., & Danoekoesoemo, W. (2016). Social Media User Personality Classification using Computational Linguistic. *International Conference on Information Technology and Electrical Engineering (ICITEE)*.

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*.

Naradhipa, A. R., & Purwarianti, A. (2012). Sentiment classification for Indonesian message in social media. *Proceedings - International Conference on Cloud Computing and Social Networking 2012: Cloud Computing and Social Networking for Smart and Productive Society, ICCCSN 2012*.

Pervin, L. A., & John, O. P. (2001). *Personality: Theory and Research (8th Edition)*. New York: John Wiley & Sons Inc.

Rumagit, R. Y., & Girsang, A. S. (2018). Predicting Personality Traits of Facebook Users Using Text Mining. *Journal of Theoretical and Applied Information Technology*, 96(20), 6877–6888.

Souri, A., Hosseinpour, S., & Rahmani, A. M. (2018). Personality Classification Based on Profles of Social Network's Users and The

Five-Factor Model of Personality. *Human-Centric Computing and Information Sciences*.

Tandera, T., Suhartono, D., Wongso, R., & Prasetio, Y. L. (2017). Personality Prediction System from Facebook Users. *2nd International Conference on Computer Science and Computational Intelligence*.