

Analisis Profil Kepribadian Aktivitas Pengguna Sosial Media

Personality Profile Analysis of Social Media User Activities

ARDIAN NUR ROMADHAN¹, EMA UTAMI², ANGGIT DWI HARTANTO³

Fakultas Magister Teknik Informatika, Universitas Amikom Yogyakarta

Email: ¹ardian.1251@students.amikom.ac.id, ²ema.u@amikom.ac.id, ³anggit@amikom.ac.id

Abstrak. Setiap individu memiliki keragaman bentuk dalam sifat dan karakter kepribadian yang dimiliki. Perbedaan karakter dari berbagai tipe kepribadian dapat dilihat dari fitur perilaku dan interaksi sosialnya, baik secara langsung maupun melalui media digital yaitu sosial media. Tentunya aktivitas dalam bentuk unggahan pengguna di sosial media dapat mencerminkan kepribadian tiap individu dari cara berkomunikasi, gaya berperilaku, dan cara menghadapi tekanan hidup. Banyaknya informasi yang dapat digali dan tersaji pada aktivitas unggahan di platform sosial media sudah dimanfaatkan oleh beberapa peneliti dalam kajian penelitian *data science* untuk klasifikasi *profiling* baik dalam bentuk gambar, audio, video atau teks. *Text mining* merupakan proses pengambilan informasi berkualitas tinggi yang bersumber dari teks. Tinjauan pustaka sistematis ini digunakan untuk menganalisis teknik penyusunan profil kepribadian aktivitas penggunaan sosial media. Dari kajian literatur ini dapat diketahui bahwa berbagai metode dan algoritma terbukti mampu melakukan prediksi klasifikasi atas kepribadian dari unggahan teks pengguna sosial media dengan tingkat akurasi yang bervariasi. Selain itu tingkat keakuratan dari setiap metode yang digunakan menunjukkan hasil bahwa diantaranya masih terdapat akurasi dibawah 50%, oleh karena itu untuk pengembangan penelitian kedepan perlu adanya peningkatan akurasi dengan pelabelan data yang tepat dan adanya peran validasi oleh Psikolog pakar di bidang Psikologi Kepribadian. Selain itu untuk penelitian mendatang juga dapat mengimplementasikan teori kepribadian, tipe kepribadian atau alat ukur kepribadian yang lain seperti 16PF (*Personality Factor Sixteen*), *The Four Temperament*, EPPS, atau MMPI.

Kata Kunci: klasifikasi kepribadian sosial media, penambangan teks, penambangan data, analitik data

Abstract. Each individual has a variety of forms in the nature and character of the personality they have. The differences in the characters of various personality types can be seen from the behavioral features and social interactions, both directly and through digital media, namely social media. Of course, activities in the form of user uploads on social media can reflect the personality of each individual from the way they communicate, behave, and how to deal with the pressures of life. The amount of information that can be extracted and presented in uploading activities on social media platforms has been used by several researchers in data science research studies for profiling classification in the form of images, audio, video or text. Text mining is the process of retrieving high-

quality information from text. This systematic literature review is used to analyze the technique of personality profiling of social media use activities. From this literature review, it can be seen that various methods and algorithms are proven to be able to predict the classification of personality from social media user text uploads with varying degrees of accuracy. In addition, the level of accuracy of each method used shows the results that some of them still have accuracy below 50%, therefore for the development of future research it is necessary to increase accuracy with proper data labeling and the role of validation by expert psychologists in the field of Personality Psychology. In addition, future research can also implement personality theory, personality types or other personality measuring tools such as 16 PF (Personality Factor Sixteen), The Four Temperament, EPPS, or MMPI.

Key words: social media personality classification, text mining, data mining, data analytics.

PENDAHULUAN

Ketertarikan akan adanya pemanfaatan *machine learning* sangat dibutuhkan di berbagai lini kehidupan, dimana salah satunya digunakan dalam menjawab pertanyaan “apakah mungkin melakukan prediksi dan pemodelan klasifikasi kepribadian berdasarkan unggahan pengguna di sosial media?”. Laman sosial media, seperti Facebook, Instagram, Twitter, Google+ dll menyediakan layanan interaksi sosial melalui media digital yang memungkinkan penggunanya saling berinteraksi di jaringan nirkabel, sehingga media tersebut menjadi sarana bertukar informasi, menyampaikan pendapat/opini atau pandangan terhadap suatu fenomena dan pengalaman maupun hal lain yang menjadi trend perhatian pada masyarakat.

Hal tersebut sering disebut dengan sentimen. Sentimen analisis merupakan suatu proses yang digunakan untuk menentukan pendapat, emosi atau sikap yang terefleksi melalui gambar, audio, video atau teks, dimana diklasifikasikan menjadi ungkapan positif atau negatif. Pemanfaatan *machine learning* dengan

menggunakan pendekatan *supervised learning* untuk melakukan analisis sentimen klasifikasi *profiling* kepribadian aktivitas pengguna sosial media saat ini menjadi daya tarik para peneliti sebagai objek penelitiannya. Keunggulan pendekatan ini dibandingkan dengan *unsupervised learning* yaitu pada pendekatan *supervised learning* terdapat algoritma yang handal dan sering digunakan dalam kajian penelitian *data science* khususnya *text mining*. Tahapan proses dalam *text mining* yaitu dari pengumpulan data, pembagian dataset, *preprocessing* teks, ekstraksi fitur hingga pemilihan algoritma dimana algoritma yang digunakan memiliki implikasi terhadap tingkat akurasi prediksi kepribadian yang dilakukan (Hartanto dkk., 2019; Thurairasa & Colombo, 2020). Berbagai penelitian klasifikasi kepribadian mencoba menganalisis kumpulan jenis data dari sosial media seseorang untuk memprediksi kepribadiannya. *Big Five*, MBTI, NEO *personality inventory*, DISC dan Ekstrovet dan Introvet merupakan beberapa teori kepribadian, tipe kepribadian atau alat ukur kepribadian yang pernah diimplementasikan oleh para peneliti terdahulu dalam kajian penelitian *data science* klasifikasi

profiling kepribadian aktivitas pengguna sosial media (lihat Tabel 1). Teori merupakan kumpulan asumsi yang saling terkait dan sistematis yang digunakan untuk menguji hipotesis. Tipe menggambarkan model dari atribut-atribut pengungkapan diri seseorang. Sedangkan alat ukur merupakan bentuk skala yang memiliki karakteristik khusus yang disusun dan dikembangkan dari variabel atau atribut, sehingga dihasilkan alat tes yang optimal, reliabel dan valid (Azwar, 2013; Farmawati & Hidayati, 2019; Fatwikiningsih, 2020; Fauzia dkk., 2019). Beberapa peneliti menggunakan metode *K-Nearest Neighbor* untuk memprediksi kepribadian pengguna sosial media (Claudy dkk., 2018; Harahap & Muslim, 2020; Munggaran & Setiawan, 2019; Putra & Wardani, 2020). Beberapa penelitian lain menggunakan metode *Support Vector Machine* untuk melakukan *profiling* kepribadian dengan memanfaatkan *text mining* (Celli & Lepri, 2018; Fahmi dkk., 2019; Fikry, 2018; Ramadhan & Putra, 2020; Soury dkk., 2018; Tighe & Cheng, 2018; Utami dkk., 2019). Selain itu berbagai metode seperti *Naïve Bayesian*, *Neural Network*, *Decision Tree*, *Logistic Regression*, *Multinomial Naïve Bayes*, *Semi-Supervised Learning*, *Gradient Boosting*, *Random Forest*, *Bayesian Optimization* juga telah digunakan untuk mengkaji pendekatan tersebut dimana penggunaan berbagai metode tersebut untuk menemukan tingkat akurasi terbaik (Claudy dkk., 2018; Hartanto dkk., 2019; Munggaran & Setiawan, 2019; Putra & Wardani, 2020;

Ramadhan & Putra, 2020; Soury dkk., 2018; Tighe & Cheng, 2018; Utami dkk., 2019, 2019; Wijaya & Cendana, 2020; Zheng & Wu, 2019). Distribusi berbagai metode yang digunakan untuk analisis *profiling* kepribadian aktivitas pengguna sosial media dapat dilihat pada Tabel 2.

Berbagai hasil penelitian yang menganalisis status unggahan di Facebook maupun analisis unggahan berupa teks/video/gambar di Instagram dan juga analisis tweet akun pengguna Twitter seseorang menunjukkan bahwa kepribadian seseorang dapat diketahui dengan menganalisis apa yang mereka unggah di sosial media baik melalui Facebook, Instagram maupun Twitter meskipun dalam proses klasifikasi menggunakan metode dan algoritma yang berbeda namun berbagai algoritma yang digunakan terbukti mampu menghasilkan tingkat akurasi prediksi yang bervariasi atas teori kepribadian, tipe kepribadian atau alat ukur kepribadian yang diimplementasikan. Dengan demikian, analisis kepribadian terhadap data unggahan berupa gambar, audio, video atau teks dari akun sosial media seseorang dapat digunakan sebagai referensi untuk mengetahui prediksi kepribadian aktivitas pengguna sosial media.

Hal ini juga menunjukkan bahwa perilaku *online* pengguna sosial media merupakan faktor yang relevan dalam memprediksikan tipe kepribadian seseorang. Berdasarkan hasil penelitian-penelitian sebelumnya terkait dengan profil kepribadian

pengguna sosial media, diperoleh hasil bahwa metode-metode yang ada pada pendekatan *supervised learning* dapat digunakan dalam pengambil prediksi kepribadian. Namun untuk memperoleh hasil lebih baik perlu adanya penelitian lebih lanjut. Kajian literatur ini akan mengulas berbagai metode yang digunakan dalam pengambilan prediksi kepribadian, sehingga dapat dimanfaatkan untuk studi penelitian kedepannya.

METODE

Metode penelitian yang digunakan untuk penulisan ini adalah tinjauan pustaka sistematis secara menyeluruh yang digunakan untuk membuat desain analisis profil kepribadian yang memanfaatkan kumpulan dataset di sosial media. Tujuan dari tinjauan ini adalah untuk mengidentifikasi diantara publikasi jurnal/artikel penelitian yang diterbitkan dengan klasifikasi *profiling* kepribadian aktivitas pengguna sosial media sebagai subjek utama atau sekundernya. Oleh karena itu, metode kerja kajian literatur ini dapat dicirikan sebagai teori dan konseptual kuantitatif.

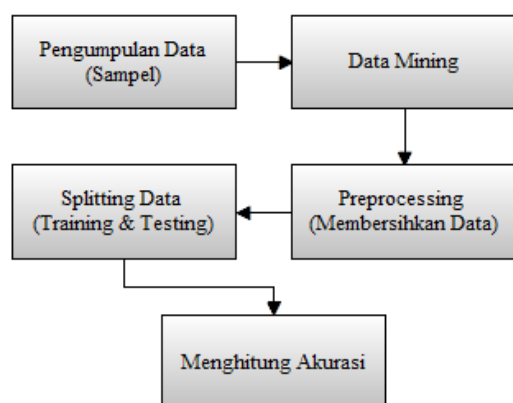
Kajian literatur ini menggunakan beberapa jenis sosial media seperti Facebook, Instagram, dan Twitter sebagai sumber datasetnya. *Screening* publikasi jurnal/artikel dilakukan antara tanggal 04 s.d 07 April 2021. Kata kunci yang digunakan untuk kajian literatur ini adalah “Klasifikasi Kepribadian *Text Mining*”. Awalnya, pencarian dilakukan

berdasarkan judul. Selanjutnya, pencarian difilter oleh bidang abstrak *text mining*. Filterisasi terhadap tahun publikasi digunakan dalam rentang 2018 s.d 2020, dengan demikian artikel yang diterbitkan kapan saja tidak bisa dimasukkan. Awalnya, 19 publikasi jurnal/artikel diidentifikasi. Namun, 4 di antaranya dibuang (*drop*), karena tidak menggunakan *text mining*. Sehingga investigasi terhadap publikasi kajian literatur ini terdiri dari lima belas publikasi jurnal/artikel. Untuk analisis data, penulis memilih untuk mempertimbangkan artikel yang diterbitkan dalam jurnal/artikel baik nasional maupun internasional. Tugas akhir, skripsi dan tesis tidak dimasukkan karena banyaknya universitas yang harus dikaji dan database akan lebih rumit. Oleh karena itu yang menjadi batasan dari kajian literatur ini, yaitu hanya menggunakan penelitian yang dipublikasikan berupa jurnal/artikel.

Kriteria klasifikasi lain yang diadopsi adalah tahun publikasi, metode algoritma, klasifikasi kepribadian dan dataset sosial media apa yang digunakan. Tujuan lainnya adalah untuk menganalisis profil kepribadian aktivitas pengguna sosial media dan mengetahui metode-metode apa saja yang dapat digunakan pada pendekatan *supervised learning* dalam melakukan prediksi kepribadian.

HASIL

Meskipun metode klasifikasi kepribadian pengguna sosial media sangat populer, bukanlah tugas yang mudah untuk mendapatkan akurasi yang baik tanpa adanya validasi dari seorang ahli yaitu Psikolog pakar di bidang Psikologi Kepribadian, dimana bidang kajian penelitian *data science* khususnya klasifikasi *profiling* kepribadian melibatkan disiplin ilmu lain yaitu Psikologi, sehingga peran dari seorang pakar ahli di bidang Psikologi Kepribadian sangat perlu dilibatkan dalam mengkonfirmasi validasi teori kepribadian, tipe kepribadian atau alat ukur kepribadian yang digunakan. Peningkatan publikasi pada topik ini sangat luar biasa selama 3 (tiga) tahun terakhir yaitu dari 2018 s.d 2020 dimana penulis hanya mengambil 5 (lima) publikasi dari tiap tahunnya. Selain itu dengan adanya penelitian yang dilakukan pada tahun 2019 terhadap topik ini, kami meyakini bahwa angka ini akan semakin meningkat dari tahun ke tahun dengan berbagai upaya perbaikan akurasi maupun perbandingan antara metode klasifikasi.



Gambar 1. Alur analisis klasifikasi *profiling* kepribadian

Analisis profil dimulai dari pengumpulan data sampel pengguna sosial media. Setelah mendapatkan sampel yang cukup, langkah selanjutnya adalah menyeleksi sampel untuk mendapatkan peserta yang valid sesuai tipe kepribadian. Setelah mendapatkan data sampel, proses selanjutnya adalah *data mining*. *Data mining* memainkan peran penting yang bertujuan untuk menggali nilai tambah dari informasi yang belum diketahui secara manual dari suatu dataset (Putra & Wardani, 2020). Dalam pengumpulan data dilakukan berbagai cara yaitu *scraping*, *crawling*, dan *steaming* (Harahap & Muslim, 2020; Utami dkk., 2019). Selanjutnya adalah proses pembersihan data yang telah difilter, proses ini biasa disebut *preprocessing*. Proses dilakukan melalui beberapa tahapan yaitu: *case folding*, *URL elimination*, *username elimination*, *hashtags elimination*, *tokenizing* atau memisahkan kalimat menjadi kata, *Normalizing words*, *Stemming*, *POS Tagging*, dan *POS Filtering*. Setelah *preprocessing*, dilakukan *splitting data* dimana pada proses ini data latih dan data pengujian dibangun. Data yang dibuat harus menyesuaikan masukan dari algoritma. Setelah data latih dan data pengujian dibuat, langkah selanjutnya adalah menguji atau menghitung tingkat akurasi. Visualisasi alur analisis klasifikasi *profiling* kepribadian dapat dilihat pada Gambar 1.

Tabel 1. Publikasi & Klasifikasi Kepribadian yang Digunakan pada Penelitian Sebelumnya

Ref. *	Peneliti	Tahun	Publikasi	Kepribadian
[01]	Claudy, Y. I., Perdana, R. S., & Fauzi, M. A.	2018	Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer	MBTI
[02]	Fikry, M.	2018	Jurnal Sains, Teknologi dan Industri	Ekstrovet atau Introvet
[03]	Souri, A., Hosseinpour, S., & Rahmani, A. M.	2018	Hum.-Centric Comput (Springer Open)	NEO <i>personality inventory</i>
[04]	Tighe, E., & Cheng, C.	2018	Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Sosial Media	<i>Big Five</i>
[05]	Celli, F., & Lepri, B.	2018	CLiC-it	<i>Big Five & MBTI</i>
[06]	Fahmi, B., Wibowo, A., & Hajar, D.	2019	Jurnal Komputer Terapan	Ekstrovet atau Introvet
[07]	Hartanto, A. D., Utami, E., Adi, S., & Hudnanto, H. S.	2019	2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE) - IEEE	DISC
[08]	Utami, E., Hartanto, A. D., Adi, S., Oyong, I., & Raharjo, S.	2019	Journal of King Saud University – Computer and Information Sciences	DISC
[09]	Munggaran, N. I. P., & Setiawan, E. B.	2019	e-Proceeding of Engineering	DISC

[10]	Zheng, H., & Wu, C.	2019	Proceedings of the 2019 11th International Conference on Machine Learning and Computing - ICMLC	<i>Big Five</i>
[11]	Idris, Utami, E., & Hartanto, A. D.	2020	TECNOSCIENZA	DISC
[12]	Wijaya, A., & Cendana, M.	2020	Jurnal Linguistik Komputasional (JLK)	MBTI
[13]	Ramadhan, M., & Putra, R. E.	2020	JINACS (Journal of Informatics and Computer Science)	<i>Big Five</i>
[14]	Putra, M. P. R., & Wardani, K. R. N.	2020	JUTIM (Jurnal Teknik Informatika Musirawas)	<i>Big Five</i>
[15]	Harahap, R. N., & Muslim, K.	2020	Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)	MBTI

*Ref = Referensi merupakan sumber acuan jurnal/artikel publikasi dimana penomoran berfungsi untuk mempertegas informasi pada Tabel 2 dan Tabel 3 atas sumber publikasi yang menjadi acuan.

Tabel 2. Dataset yang Digunakan pada Penelitian Sebelumnya

Ref.	Metode Algoritma	Dataset	Akurasi
[01]	<i>K-Nearest Neighbor</i>	Twitter	66%
[02]	<i>Support Vector Machine</i>	Twitter	88,89%.
[03]	<i>Naïve Bayesian, Neural Network, Decision Tree, Support Vector Machine</i>	Facebook	NB 67.46%; NN 86.67%; DT 93.33%; SVM 95.56%
[04]	<i>Support Vector Machine, Logistic Regression</i>	Twitter	F1 SVM:0.5669 F1

			LOG: 0.6086
[05]	<i>Support Vector Machine</i>	Twitter	Big5: 61.5%; MBTI: 64.6%
[06]	<i>Support Vector Machine</i>	Facebook	83.33%.
[07]	<i>Naïve Bayes Classifier</i>	Twitter	36.67%.
[08]	<i>Support Vector Machine</i>	Twitter	37,41%, 30,21%, 35,97%, dan 30,93%
[09]	<i>K-Nearest Neighbor</i>	Twitter	40.60%
[10]	<i>Semi-Supervised Learning</i>	Facebook & Twitter	F1: 0.71 Precision: 0.69 Recall: 0.73
[11]	<i>Artificial Neural Network</i>	Twitter	42,91%
[12]	<i>Naive Bayes Classifier</i>	Twitter	71%
[13]	<i>Support Vector Machine dan Bayesian Optimization</i>	Instagram	Akurasi 60.34%, presisi 30.17%, dan recall 50%
[14]	<i>K-Nearest Neighbor</i>	Facebook	92.92%
[15]	<i>Gradient Boosting, Multinomial Naïve Bayes, Random Forest, K-Nearest Neighbor</i>	Twitter	GB 30%; MNB 20%; RF 40%; KNN 20%

Tabel 3. Analisis Perbandingan Hasil Ulasan Penelitian Sebelumnya

Ref.	Tujuan Penelitian	Kesimpulan	Saran / Kelemahan
[01]	Untuk mengetahui karakter calon karyawan berdasarkan tweet yang diolah untuk menghasilkan ciri	Hasil kajian penelitian ini menghasilkan nilai akurasi sebesar 66% dengan 53 data terklasifikasi benar dan 27 data	Penelitian selanjutnya dapat dilanjutkan dengan menggunakan metode pengujian akurasi lainnya,

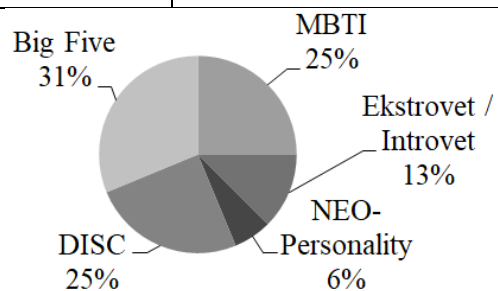
	kepribadian MBTI.	salah klasifikasi	seperti k-fold cv.
[02]	Untuk dapat memahami kepribadian seseorang, profil sosial medianya dapat digunakan sebagai sumber informasi.	Dari hasil pengujian diketahui bahwa metode SVM dapat diterapkan untuk mengklasifikasi kepribadian Ekstrovet dan introvet dengan akurasi sebesar 88,89%.	Penelitian ini dapat dilanjutkan dengan menggunakan fitur-fitur lain yang belum digunakan. Fitur-fitur tersebut memerlukan data tambahan dari akun lain yang menyebutkan (mention), me-retweet, menyukai status (favorit) dari si pemilik akun.
[03]	Penelitian ini bertujuan untuk mengenali kepribadian Lima Faktor seseorang berdasarkan analisis aktivitas penggunaannya di Facebook	Model yang diusulkan dengan akurasi 82,2% lebih akurat daripada penelitian sebelumnya	Penelitian selanjutnya diharapkan dapat menggunakan peserta yang berasal dari negara berbeda dan kondisi berbeda.
[04]	Untuk membuat model kepribadian berbasis teks tentang bagaimana orang Filipina berbicara.	Analisis representasi skor ciri kepribadian menunjukkan bahwa pengklasifikasian pencilan ekstrim umumnya menghasilkan hasil yang lebih baik untuk semua sifat kecuali untuk Neurotisme dan Keterbukaan.	Mengumpulkan data dari pengguna yang sama yang dianotasi dengan MBTI dan <i>Big Five</i> dengan kueri Twitter, ini adalah sesuatu yang dapat dilakukan hanya dengan upaya pengumpulan data yang mahal, yang kami harap dapat dilakukan di masa mendatang.
[05]	Perbandingan dari dua model <i>Big Five</i> dan MBTI dari perspektif komputasi.	Algoritma yang dilatih pada MBTI dapat memiliki kinerja yang lebih baik daripada yang dilatih pada Big5.	Dalam kinerja tergantung juga pada klasifikasi algoritma yang digunakan untuk prediksi.

[06]	Untuk membuat sebuah sistem yang dapat menganalisa kepribadian pengguna Facebook berdasarkan status yang pernah diunggahnya.	Dari 158 data status Facebook didapatkan hasil pengujian sebesar 83,3% dari hasil perbandingan antara pengeluaran sistem yang dibuat dengan sistem yang telah ada.	Penelitian selanjutnya dapat menggunakan TF-IDF atau W-IDF untuk pembobotan kata.
[07]	Untuk mendapatkan data kepribadian seseorang melalui tweet dalam pencarian SDM.	Klasifikasi data tweet Dominasi 90 akun, Pengaruh 10 akun, Steadiness 8 akun dan Kepatuhan 12 akun.	Penelitian selanjutnya perlu memperbanyak jumlah data pelatihan kata yang telah diberi label serta akun Twitter yang telah divalidasi oleh psikolog.
[08]	Untuk dapat mengamati dan menilai kepribadian pengguna dari pilihan kata di tweet mereka, kami membuat kosakata kata kunci yang relevan sesuai dengan kerangka kerja dan teori DISC.	Keempat skenario yang menggunakan kosakata kata kunci tidak bertangkai-tidak berbobot, bertangkai-tidak berbobot, tidak bertangkai, dan berbobot-berbobot, dan tingkat akurasi masing-masing 30,93%	Penelitian selanjutnya dapat menggunakan Chi-Square sebagai pendekatan yang sesuai untuk proses pemilihan fitur.
[09]	Melihat seberapa akurat penggunaan algoritma KNN dalam pengklasifikasian kepribadian pengguna twitter serta seberapa besar perbedaan antara 2 metode pembobotan.	Pembobotan dengan TF-IDF dan TF-Chi Square tidak terlalu mempengaruhi akurasi terbukti dengan nilai akurasi pendekatan perilaku sosial lebih baik dari pada linguistik.	Penelitian selanjutnya mencari algoritma yang bisa menangani masalah ketimpangan data agar hasil agar keputusan model yang dibuat tidak cenderung ke kelas yang paling dominan.
[10]	Untuk mengeksplorasi penggunaan data tak berlabel pada analisis kepribadian, kerangka analisis kepribadian	Hasil eksperimen menunjukkan bahwa pembelajaran <i>semi-supervised</i> dapat memanfaatkan data tidak	Karena banyak penelitian telah menemukan keterkaitan antara konten buatan pengguna dengan

	berdasarkan pembelajaran semi-supervisi.	berlabel dan meningkatkan akurasi model prediksi.	ciri-ciri kepribadian, beberapa pengguna sosial media menggunakan langkah perlindungan privasi terhadap akunnya.
[11]	Untuk mengetahui seberapa besar akurasi analisis <i>profiling</i> pada Twitter sehingga bisa menjadi acuan untuk proses perekrutan pegawai	Algoritma Artificial Neural Network dapat menghasilkan akurasi terbaik sebesar 42.91%	Penelitian selanjutnya dapat menggunakan variabel-variabel lain selain teks, seperti emoji, gambar, tagar, lokasi, dan lain-lain agar dapat meningkatkan akurasi
[12]	Merancang aplikasi berbasis web yang dapat melakukan klasifikasi tipe kepribadian seseorang berdasarkan cuitan di Twitter diharapkan dapat menjadi salah satu alternatif untuk mengetahui tipe kepribadian seseorang tanpa perlu mengikuti psikotes secara konvensional	Tingkat akurasi sebesar 71% dengan rincian precision dan recall untuk kelas artisan sebesar 63% dan 68%, untuk kelas guardian sebesar 77% dan 68%, untuk kelas rational sebesar 64% dan 72%, dan untuk kelas idealist sebesar 83% dan 76%.	Penelitian selanjutnya teknik pra pemrosesan dapat diperkaya dengan penambahan stopword, kombinasi stemming dan lemmatization.
[13]	Untuk menerapkan algoritma <i>Support Vector Machine</i> (SVM) pada sistem prediksi kepribadian pengguna Instagram	Kombinasi algoritma <i>Support Vector Machine</i> dengan kernel RBF dan Bayesian Optimization mampu menghasilkan akurasi mencapai 60.34%, presisi 30.17%, dan recall 50% untuk mengklasifikasi kepribadian pengguna Instagram.	Penelitian selanjutnya dapat membandingkan tingkat akurasi yang dihasilkan oleh model SVM dengan model klasifikasi pembanding lainnya.
[14]	Untuk mengetahui informasi apa”saja”yang dapat diambil	Tingkat akurasi yang diperoleh yaitu 92.92%, dari 550 data	Penelitian ini masih memiliki kelemahan dalam

	tentang kepribadian pengguna.	dengan nilai karakter kepribadian openness yang paling tinggi yaitu 239, Conscientiousness sebanyak 16 data, Extraversion sebanyak 173 data, Agreeableness sebanyak 50 data, Neuroticism sebanyak 33 data dan 39 data yang tidak dapat diklasifikasikan.	mencocokkan kepribadian <i>Big Five Personality</i> .
[15]	Penelitian ini mencoba memberikan solusi dengan membangun model yang dapat melakukan prediksi terhadap kepribadian seseorang berdasarkan analisis data dan informasi dari sosial media Twitter.	Hasil dari Penelitian ini menunjukkan bahwa teknik augmentasi data cukup efektif dilakukan untuk meningkatkan performa <i>text mining</i> task dengan jumlah dataset sedikit. Hal ini terlihat pada peningkatan akurasi dari model prediksi kepribadian MBTI yang dibangun.	Penelitian selanjutnya diharapkan menambah jumlah dataset dan jumlah tweet yang seimbang pada setiap kelas untuk mendapatkan akurasi yang lebih baik.

Tipe kepribadian merupakan komponen yang dapat mempengaruhi pengungkapan diri seorang individu terutama ketika sedang berinteraksi digital menggunakan media sosial (Fauzia dkk., 2019). Pada kajian literatur ini terdapat 5 (lima) klasifikasi kepribadian berupa teori kepribadian, tipe kepribadian dan alat ukur kepribadian yang terdistribusi yaitu *Big Five*, MBTI, *NEO personality inventory*, DISC dan Ekstrovet dan Introvet (lihat Tabel 1).

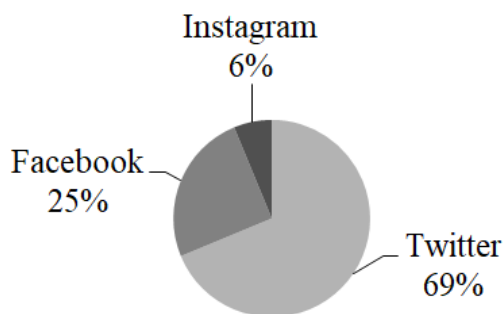


Gambar 2. Distribusi Klasifikasi Kepribadian

Klasifikasi kepribadian merupakan salah satu tantangan baru dalam kajian penelitian *data science* pada klasifikasi *profiling* dengan menggunakan *platform* sosial media. Unggahan berupa audio, video, teks, dan gambar diharapkan menampilkan pola

perilaku kepribadian saat pengguna berinteraksi di sosial media. *Big Five* dengan persentase 31% merupakan kepribadian paling populer yang digunakan sebagai klasifikasi pada kajian literatur ini (lihat Gambar 2) diikuti DISC dan MBTI dimana masing-masing memiliki persentase sebesar 25%. Sedangkan klasifikasi kepribadian lainnya berada di kisaran persentase 6% s.d 13%.

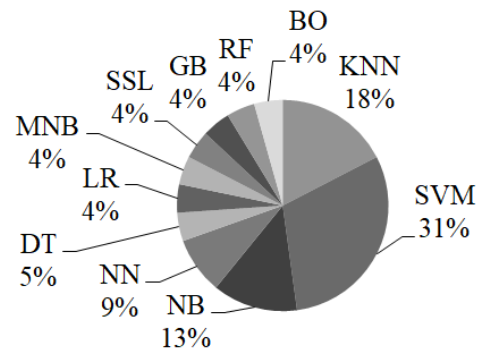
Sosial media atau jejaring *online* telah menjadi cara yang dituntut bagi pengguna untuk menunjukkan eksistensi diri. Selain dimanfaatkan untuk berbagi interaksi digital, sosial media juga seringkali difungsikan sebagai sarana pengungkapan diri (Fauzia dkk., 2019). Pada *platform* sosial media tersebut berbagi informasi dan komunikasi di antara pengguna sosial media dapat diambil sebagai data. Pada kajian literatur ini terdapat 3 (tiga) sosial media yang terdistribusi yaitu Facebook, Instagram, dan Twitter (lihat Tabel 2).



Gambar 3. Distribusi Jejaring Sosial Media

Twitter dengan persentase 69% merupakan jejaring sosial media paling populer yang digunakan sebagai dataset pada kajian literatur ini (lihat Gambar 3) diikuti Facebook dengan persentase sebesar 25% dan

Instagram dengan persentase 6%. Kebanyakan peneliti klasifikasi kepribadian pengguna sosial media menggunakan dataset dari twitter (lihat Tabel 1). Hal ini dimungkinkan karena twitter merupakan sosial media berbasis *text*, sehingga sangat cocok terhadap penelitian yang berbasis *text mining*.



Gambar 4. Distribusi Metode Algoritma

Distribusi dari metode algoritma pada kajian literatur ini terdapat 11 (sebelas) metode yaitu *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), *Naïve Bayesian* (NB), *Neural Network* (NN), *Decision Tree* (DT), *Logistic Regression* (LR), *Multinomial Naïve Bayes* (MNB), *Semi-Supervised Learning* (SSL), *Gradient Boosting* (GB), *Random Forest* (RF), *Bayesian Optimization* (BO) (lihat Tabel 2).

SVM dengan persentase 31% merupakan pendekatan *supervised learning* yang paling populer atau banyak digunakan sebagai metode algoritma untuk melakukan *profiling* kepribadian pengguna sosial media pada kajian literatur ini (lihat Gambar 4). Diikuti KNN 18% dan NB memiliki persentase sebesar 13%. Sedangkan metode algoritma

lainnya berada di kisaran persentase 4% s.d 9%.

Akurasi dari masing-masing algoritma yang digunakan dalam proses klasifikasi *profiling* atas berbagai teori kepribadian, tipe kepribadian atau alat ukur kepribadian menunjukkan tingkat akurasi yang bervariasi. Pada Tabel 2 dapat diketahui bahwa beberapa metode algoritma yang diimplementasikan mampu merepresentasikan teori kepribadian, tipe kepribadian atau alat ukur kepribadian dengan persentase akurasi mendekati 100%, namun beberapa diantaranya memiliki akurasi dibawah 50% hal tersebut menunjukkan bahwa perlu adanya perbaikan pada penelitian mendatang untuk peningkatan akurasi, sehingga algoritma yang digunakan dapat merepresentasikan atau melakukan prediksi atas teori kepribadian, tipe kepribadian atau alat ukur kepribadian yang digunakan dalam klasifikasi *profiling* kepribadian aktivitas pengguna sosial media.

Semua hasil distribusi disajikan secara lebih rinci di Tabel 1, Tabel 2, dan Tabel 3 dimana terdapat klasifikasi masing-masing dari 15 jurnal/artikel. Dalam tabel yang sama juga terdapat jurnal/artikel yang sesuai untuk setiap dataset, tahun publikasi, klasifikasi kepribadian dan algoritma *machine learning supervised* yang digunakan.

SIMPULAN

Dari kajian literatur ini dapat diketahui bahwa berbagai metode dan algoritma dapat

melakukan prediksi klasifikasi atas kepribadian dari unggahan teks pengguna sosial media dengan tingkat akurasi yang bervariasi. Dapat dilihat pada Gambar 3 yang ada pada kajian literatur ini bahwa implementasi teori kepribadian, tipe kepribadian atau alat ukur kepribadian yang populer atau banyak digunakan yaitu *Big Five* dan DISC tentunya untuk teori kepribadian, tipe kepribadian atau alat ukur kepribadian lain juga dapat dijadikan alternatif dalam memvisualisasikan kepribadian pengguna sosial media sesuai kebutuhan yang diinginkan oleh pihak berkepentingan, sedangkan pada Gambar 4 metode KNN, SVM, dan NB merupakan pilihan yang paling banyak digunakan dalam implementasi *profiling* aktivitas pengguna sosial media. Namun dari keseluruhan metode baik KNN, SVM, NB, NN, DT, LR, MNB, SSL, GB, RF, dan BO semuanya dapat mengatasi prediksi kepribadian.

Hasil analisis kepribadian tersebut nantinya dapat dimanfaatkan oleh pihak berkepentingan. Sebagai contoh dalam proses seleksi dan rekrutmen karyawan disebuah perusahaan. Dengan demikian, dapat disimpulkan bahwa analisis *profiling* melalui aktivitas pengguna sosial media dapat digunakan sebagai salah satu acuan atau referensi untuk mengetahui prediksi kepribadian pengguna sosial.

DISKUSI

Kajian literatur ini menganalisis *profiling* melalui aktivitas pengguna platform sosial media. Data informasi aktivitas tentang seseorang dapat digali melalui unggahannya di sosial media seperti Facebook, Instagram, dan Twitter. Unggahan di sosial media dapat dijadikan sebagai sumber data untuk analisis, sehingga dapat diperoleh informasi penting tentang kepribadian seorang individu. Analisis *profiling* ini didiskusikan untuk mengembangkan skala penelitian yang lebih besar di bidang ini.

Saran/kelemahan penelitian terdahulu atas 15 jurnal/artikel pada kajian literatur ini tersaji dalam Tabel 3. Akhirnya kajian literatur ini memberikan saran untuk penelitian mendatang agar memperhatikan perlunya melakukan perbandingan antar tiap metode algoritma yang digunakan, sehingga dapat diketahui tingkat akurasi dari tiap masing-masing metode apakah sama atau ada yang lebih unggul dan handal dalam proses klasifikasi kepribadian aktivitas pengguna sosial media. Selain itu untuk tingkat akurasi dari beberapa metode masih terdapat akurasi dibawah 50%, oleh karena itu untuk pengembangan penelitian kedepan perlu adanya peningkatan akurasi dengan pelabelan data yang tepat dan adanya peran validasi oleh Psikolog pakar di bidang Psikologi Kepribadian dan untuk penelitian mendatang juga dapat menggunakan variabel lain selain teks, seperti gambar, audio, atau video. Selain

dari pada hal tersebut untuk penelitian mendatang dapat menerapkan teori kepribadian, tipe kepribadian atau alat ukur kepribadian yang lain seperti 16 PF (*Personality Factor Sixteen*), *The Four Temperament*, EPPS (*Edwards Personal Preference Schedule*), atau MMPI (*Minnesota Multiphasic Personality Inventory*).

DAFTAR PUSTAKA

- Azwar, S. (2013). *Penyusunan skala psikologi (Edisi 2)*. Yogyakarta: Pustaka Pelajar.
- Celli, F., & Lepri, B. (2018). *Is Big Five better than MBTI? A personality computing challenge using Twitter data*. 93–98. <https://doi.org/DOI:10.4000/books.aacademia.3147>
- Claudy, Y. I., Perdana, R. S., & Fauzi, M. A. (2018). *Klasifikasi Dokumen Twitter Untuk Mengetahui Karakter Calon Karyawan Menggunakan Algoritma K-Nearest Neighbor (KNN)*. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(8), 2761–2765.
- Fahmi, B., Wibowo, A., & Hajar, D. (2019). *Analisa Kepribadian Pengguna Facebook Menggunakan Algoritma Support Vector Machine*. *Jurnal*

- Komputer Terapan*, 5(1), 28–35.
<https://doi.org/10.35143/jkt.v5i1.2259>
- Farmawati, C., & Hidayati, N. (2019). *Penyusunan dan pengembangan alat ukur Islamic personality scale (IPS)*. *Jurnal Psikologi Islam dan Budaya*, 2(1), 19–30.
- Fatwikiningsih, N. (2020). *Teori Psikologi Kepribadian Manusia*. Yogyakarta: Penerbit Andi.
- Fauzia, A. Z., Maslihah, S., & Ihsan, H. (2019). *Pengaruh tipe kepribadian terhadap self-disclosure pada dewasa awal pengguna media sosial instagram di kota bandung*. *Journal of Psychological Science and Profession*, 3(3), 151–160.
- Fikry, M. (2018). *Ekstrover atau Introver: Klasifikasi Kepribadian Pengguna Twitter dengan Menggunakan Metode Support Vector Machine*. *Jurnal Sains, Teknologi dan Industri*, 16(1), 72.
<https://doi.org/10.24014/sitekin.v16i1.5326>
- Harahap, R. N., & Muslim, K. (2020). *Peningkatan Akurasi pada Prediksi Kepribadian Mbti Pengguna Twitter Menggunakan Augmentasi Data*. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(4), 815.
<https://doi.org/10.25126/jtiik.2020743622>
- Hartanto, A. D., Utami, E., Adi, S., & Hudnanto, H. S. (2019). *Job seeker profile classification of twitter data using the naïve bayes classifier algorithm based on the DISC method*. *4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 533–536.
- Munggaran, N. I. P., & Setiawan, E. B. (2019). *Prediksi Kepribadian DISC dengan K-Nearest Neighbour Algorithm (KNN) Menggunakan Pembobotan TF-IDF dan TF-Chi Square*. *e-Proceeding of Engineering*, 6(7), 9446–9457.
- Putra, M. P. R., & Wardani, K. R. N. (2020). *Penerapan Text Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial*. *JUTIM (Jurnal Teknik Informatika Musirawas)*, 5(1), 63–71.
<https://doi.org/10.32767/jutim.v5i1.791>
- Ramadhan, M., & Putra, R. E. (2020). *Prediksi Kepribadian Pengguna Instagram*

- Berdasarkan Model Big Five Personality Menggunakan Algoritma SVM. JINACS: Journal of Informatics and Computer Science*, 01(4), 179–187.
- Souri, A., Hosseinpour, S., & Rahmani, A. M. (2018). *Personality classification based on profiles of social networks' users and the five-factor model of personality. Human-Centric Computing and Information Sciences*, 8(1), 24.
- Thurairasa, T., & Colombo, S. (2020). *Personality Prediction Review on Text Modality Dataset. Information Technology Research*, 57–61.
- Tighe, E., & Cheng, C. (2018). *Modeling personality traits of Filipino twitter users. In Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, 112–122.
- Utami, E., Hartanto, A. D., Adi, S., Oyong, I., & Raharjo, S. (2019). *Profiling analysis of DISC personality traits based on Twitter posts in Bahasa Indonesia. Journal of King Saud University - Computer and Information Sciences*, S131915781930744X.
- Wijaya, A., & Cendana, M. (2020). *Klasifikasi Kepribadian Myres-Briggs Type Indicator Berdasarkan Cuitan di Twitter Menggunakan Metode TF-IDF dan Naive Bayes Classifier. Jurnal Linguistik Komputasional*, 3(2), 48–53.
- Zheng, H., & Wu, C. (2019). *Predicting Personality Using Facebook Status Based on Semi-supervised Learning. Proceedings of the 2019 11th International Conference on Machine Learning and Computing - ICMLC '19*, 59–64.