

Ekstrover atau Introver : Klasifikasi Kepribadian Pengguna Twitter dengan Menggunakan Metode Support Vector Machine

Muhammad Fikry¹, Yusra²

^{1,2} Jurusan Teknik Informatika Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau
Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, 28293
Email: muhammad.fikry@uin-suska.ac.id, yusra@uin-suska.ac.id

ABSTRAK

Orang bertipe kepribadian ekstrover dan introver menggunakan media sosial namun dalam tingkatan berbeda dan dengan alasan yang berbeda pula. Untuk dapat memahami kepribadian seseorang, profil media sosialnya dapat digunakan sebagai sumber informasi. Pada penelitian ini, dilakukan klasifikasi kepribadian pengguna Twitter ke dalam kelas ekstrover atau introver dengan menggunakan metode Support Vector Machine (SVM). Profil pengguna Twitter diunduh melalui Twitter API sebanyak 46 akun. Pelabelan tipe kepribadian dilakukan berdasarkan hasil kuesioner kepribadian. Jumlah ciri yang digunakan sebanyak 17 ciri. Fungsi *kernel* yang digunakan adalah *kernel* RBF. Setelah melakukan seleksi fitur dan seleksi parameter, dilakukan pelatihan untuk mendapatkan model terbaik. Model tersebut digunakan terhadap data uji. Dari hasil pengujian diketahui bahwa metode SVM dapat diterapkan untuk mengklasifikasi kepribadian ekstrover dan introver dengan akurasi sebesar 88,89%.

Kata Kunci: klasifikasi, ekstrover, introver, support vector machine, tweet

ABSTRACT

People with extraverted and introverted personality type use social media in different ways for different reasons. In order to understand people's personality, their social media profiles can be used as a source of information. In this research, Support Vector Machine method is used to classify Twitter user's personality as an extrovert or introvert. Twitter user's profile of 46 accounts are downloaded through the Twitter API. Labeling of personality types is based on the results of personality questionnaires. Number of features used is 17 features. RBF kernel is used as kernel function. After performing feature selection and parameter selection, we train the training dataset to get the best model. The model is used to predict the test data. From the test results, SVM method can be used to classify Twitter user's personality with 88,89% accuracy.

Keywords: classification, extrovert, introvert, support vector machine, tweet

Corresponding author:

Muhammad Fikry

Program Studi Teknik Informatika, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau,
Email: muhammad.fikry@uin-suska.ac.id

Pendahuluan

Tipe kepribadian ekstrover (*extraverted type*) dan introver (*introverted type*) dipopulerkan oleh Jung [1]. Kepribadian ekstrover lebih menyukai interaksi dengan orang lain, antusias dalam hal baru, senang bergaul dan senang menjadi pusat perhatian. Sebaliknya, kepribadian introver cenderung menyukai ketenangan, kesendirian, dan reflektif terhadap hal-hal yang telah dilakukannya. Kamus Besar Bahasa Indonesia (KBBI) mendefinisikan ekstrover sebagai sikap terbuka yang minatnya ditunjukkan seluruhnya kepada yang ada di luar dirinya. Sementara itu, introver didefinisikan sebagai sifat tertutup yang suka

memendam rasa dan pikiran sendiri, dan tidak mengutarakannya kepada orang lain.

Setiap orang memiliki sisi ekstrover dan introver, dimana salah satu sisi lebih dominan dari lainnya. Tingkat ekstrover dan introver biasanya dinilai melalui pengukuran *self-report*. Selain itu, *peer-report* dan observasi pihak ketiga juga dapat digunakan.

Dengan mengetahui tipe kepribadian seseorang, hal ini akan membantu dalam menerima dirinya (*self-acceptance*) dan memahami orang lain yang memiliki tipe kepribadian berbeda. Salah satunya adalah memahami adanya perbedaan gaya komunikasi antara orang bertipe ekstrover dengan orang bertipe introver [2].

Situs jejaring sosial, seperti Facebook, Twitter dan LinkedIn, menyediakan layanan yang memungkinkan penggunanya saling berkomunikasi di dunia maya dalam bentuk pesan. Untuk mengetahui peranan tipe kepribadian ekstrover dan introver dalam berkomunikasi di media sosial, Schaubhut, Weber dan Thompson [3] melakukan survey terhadap 1784 orang pengguna media sosial. Berdasarkan hasil survey, diketahui bahwa tipe kepribadian berperan penting. Ekstrover dan introver sama-sama menggunakan media sosial namun dalam tingkatan berbeda dan dengan alasan yang berbeda pula.

Orang bertipe ekstrover jauh lebih banyak terlibat dalam interaksi sosial dibandingkan orang bertipe introver [3]. Hal ini dikarenakan tingkat ekstrover memiliki pengaruh terbesar pada aktifitas seseorang di media sosial [4]. Menurut Widiyanti dan Herdiyanto [5], ekstrover memiliki intensitas komunikasi lebih tinggi dibandingkan introver. Namun demikian, tidak terdapat perbedaan intensitas komunikasi melalui media sosial antara laki-laki dan perempuan. Hasil-hasil penelitian ini konsisten dengan kecenderungan ekstrover untuk lebih senang berkomunikasi dengan lingkungan di luar dirinya.

Menurut Zywica dan Danowski [6], fenomena tersebut di atas dikenal sebagai *social enhancement hypothesis*, yang menyatakan bahwa individu ekstrover dan terbuka akan termotivasi untuk memperbanyak kontak *online*. Hal ini juga memperlihatkan bahwa tipe kepribadian merupakan faktor yang relevan dalam menentukan perilaku *online*.

Berbeda dengan ekstrover yang memandang media sosial sebagai perpanjangan dalam interaksi sosial, orang bertipe introver memandang media sosial sebagai pengganti interaksi berhadapan-hadapan [7]. Introver dapat memilih dan memilah siapa yang menjadi temannya. Introver juga dapat mengontrol banyaknya interaksi sosial yang dihadapinya. Dengan demikian, introver dapat dengan mudah mengekspresikan dirinya tanpa perlu menyampaikannya secara berhadapan-hadapan.

Setelah pesan dibuat, tentunya lebih banyak balasan komentar yang berasal dari pendengar yang ekstrover. Namun demikian, introver juga mendengarkan meskipun tak terlihat. Oleh karena itu, suatu pesan perlu disesuaikan dengan tipe kepribadian pendengarnya sehingga penyampaiannya menjadi lebih efektif. Untuk dapat melakukan hal tersebut, terlebih dahulu perlu diketahui tipe kepribadiannya secara otomatis, tanpa perlu melakukan pengukuran *self-report* yang melibatkan pengisian kuesioner dan psikolog.

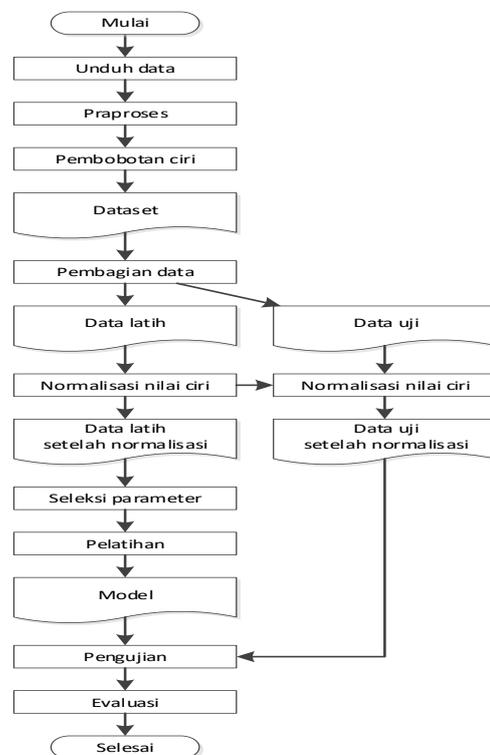
Profil media sosial dapat digunakan sebagai sumber informasi untuk memahami seseorang. Golbeck, Robles dan Turner [8] memprediksi kepribadian pengguna Facebook berdasarkan informasi yang tersedia untuk umum di profil Facebook mereka. Golbeck, dkk. [9] juga memprediksi kepribadian (Big Five Personality) pengguna Twitter. Profil pengguna Twitter dan *tweet*-nya juga digunakan oleh Qiu, dkk. [10] untuk memprediksi ciri kepribadian Big Five

(*openness, conscientiousness, extraversion, agreeableness, neuroticism*), Sarwani dan Mahmudy [11] untuk mengklasifikasikan empat temperamen dasar (*artisan, guardian, idealist rational*), serta Ahmad dan Siddique [12] untuk mengklasifikasikan empat model kepribadian DISC (*dominance, influence, steadiness, compliance*). Untuk tipe kepribadian ekstrover dan introver, Nguyen, dkk. [13] melakukan klasifikasi pengguna LiveJournal berdasarkan fitur hubungan sosial (jumlah *follower*, jumlah *friend*, dan jumlah keanggotaan komunitas), psikolinguistik (perilaku berbahasa) dan *mood* yang berasal dari pesan teks. Metode SVM yang digunakannya mendapatkan hasil terbaik sebesar 80% dengan memanfaatkan fitur psikolinguistik.

Dalam penelitian ini, diklasifikasikan kepribadian pengguna Twitter ke dalam kelas Ekstrover atau Introver, dengan menggunakan metode Support Vector Machine [14], dengan memanfaatkan fitur berbeda dibandingkan dengan [13]. SVM bertujuan menemukan *hyperplane* yang dapat memisahkan kumpulan data ke dalam dua kelas berbeda dengan margin terbesar. Margin merupakan jarak antara *hyperplane* dengan data terdekat pada masing-masing kelas yang disebut *support vector*. SVM mengklasifikasi data yang belum diketahui kelasnya berdasarkan pada sisi mana dari *hyperplane* data tersebut muncul.

Metode Penelitian

Tahapannya diperlihatkan pada Gambar 1.



Gambar 1. Tahapan penelitian

Ciri (fitur) yang digunakan dalam penelitian ini berdasarkan penelitian-penelitian berikut dengan penyesuaian terhadap media sosial Twitter. Ciri ke-1 s.d. 10 mengacu kepada Riquelme dan Gonzalez-Cantergiani [15], dengan penyesuaian pada ciri ke-4 dan 10, serta penambahan ciri ke-11 dan 12 berdasarkan Pal dan Counts [16]. Samghabadi, dkk. [17] menjadi dasar untuk ciri lainnya.

Ciri-ciri selengkapnya sebagai berikut :

1. Jumlah *tweet* yang dibuat oleh pemilik akun. *Tweet* tersebut tidak diawali *mention*, tidak mengandung simbol RT dan bukan *tweet* balasan.
2. Jumlah URL yang ditemukan pada ciri ke-1.
3. Jumlah *hashtag* yang ditemukan pada ciri ke-1.
4. Jumlah *retweet* dan *quoted tweet*, mengandung simbol RT.
5. Jumlah *tweet* pengguna lain yang disukai oleh pemilik akun, dikenal sebagai *liked* atau *favorite*.
6. Jumlah *mention* oleh pemilik akun.
7. Jumlah pengguna unik yang di-*mention* oleh pemilik akun.
8. Jumlah *follower* yaitu akun-akun yang mengikuti pemilik akun.
9. Jumlah *following* yaitu akun-akun yang diikuti oleh pemilik akun. Lebih dikenal dengan istilah teman.
10. Keaktifan si pemilik akun berdasarkan aksi yang terlihat, seperti *tweet*, *retweet*, *mention* dan *reply*. Nilainya adalah hasil penjumlahan ciri ke-1, 4, 5, 11 dan 12.
11. Jumlah *tweet* percakapan yang dimulai oleh pemilik akun, diawali *mention*, tidak mengandung simbol RT.
12. Jumlah *tweet* balasan (*reply*) kepada pengguna lain. *Tweet* tersebut merupakan *tweet* percakapan, namun tidak dimulai oleh pemilik akun.
13. Jumlah kata pada deskripsi profil.
14. Rata-rata jumlah kata per *tweet*.
15. Rata-rata jumlah karakter per *tweet*.
16. Jumlah *emoticon* dan *emoji*.
17. Jumlah media (foto dan video).

Berdasarkan data penelitian oleh Hamdila [18], diperoleh sebanyak 60 akun Twitter berbahasa Indonesia beserta tipe kepribadiannya. Setiap pemilik akun telah mengisi kuesioner kepribadian ekstrover dan introver, kemudian ditentukan tipe kepribadiannya oleh seorang psikolog. Oleh karena sebagian besar ciri yang digunakan berbeda, dilakukan pengumpulan ulang profil beserta *tweet*-nya melalui Twitter API. Selain itu, jumlah maksimal *tweet* yang diunduh dinaikkan mencapai batas 3200 *tweet* terbaru, jauh lebih banyak dari sebelumnya yaitu 200 *tweet*.

Data akun yang terkumpul dipraproses dan dicari nilai setiap ciri yang digunakan. Dengan demikian, setiap data mengandung satu nilai target yaitu label kelas bernilai tipe kepribadian, dan sejumlah atribut yaitu ciri beserta nilainya.

Klasifikasi dilakukan dengan menggunakan LibSVM [19]. Fungsi *kernel* yang digunakan adalah fungsi *kernel* yang umum digunakan, yaitu *gaussian radial basis function* (RBF) dengan pasangan parameter C dan γ .

Sebelum dilakukan klasifikasi, data akun dibagi menjadi kumpulan data latih (*training*) dan kumpulan data uji (*testing*). Selanjutnya, dilakukan *scaling* (penskalaan) berupa normalisasi nilai ciri ke dalam rentang tertentu. Hal ini dimaksudkan untuk menghindari atribut dalam rentang numerik yang lebih besar mendominasi atribut dalam rentang numerik yang lebih kecil.

Untuk mendapatkan pasangan parameter terbaik, dilakukan *grid search* dan *cross validation* untuk setiap kemungkinan pasangan parameter. Pasangan parameter dengan akurasi *cross validation* tertinggi yang akan digunakan untuk melatih seluruh data latih dan menghasilkan model. Model inilah yang digunakan untuk memprediksi label kelas dari data uji. Hasil akurasi pengujian dapat dirumuskan sebagai berikut.

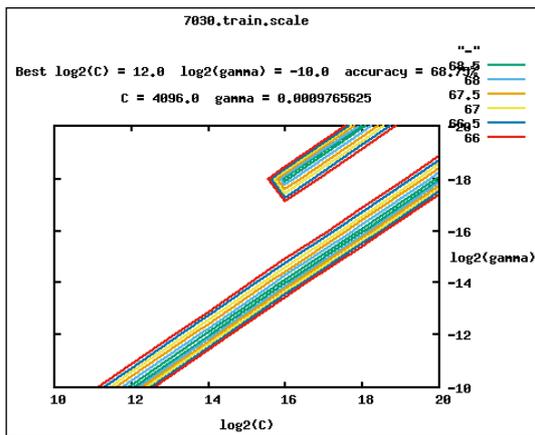
$$\text{Akurasi} = \frac{\text{Jumlah klasifikasi benar}}{\text{Jumlah dokumen}} \times 100\% \quad (1)$$

Hasil dan Pembahasan

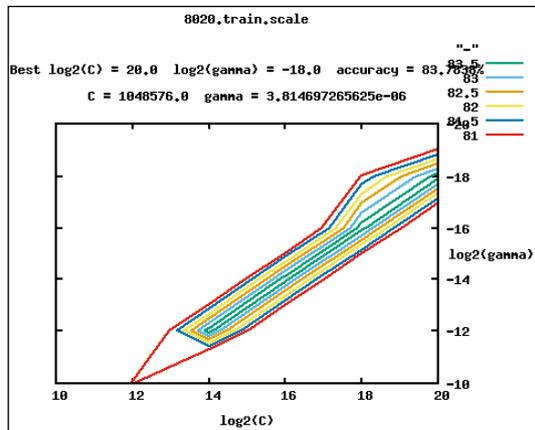
Pengumpulan ulang profil hanya menghasilkan sebanyak 51 akun Twitter, sementara sisanya sebanyak 9 akun tidak dapat diakses oleh publik karena telah dilindungi (*protected*) atau dihapus pada saat penelitian ini dilakukan. Untuk memperoleh jumlah data akun yang berimbang pada setiap label kelasnya, dipilih sebanyak 23 akun bertipe ekstrover dan 23 akun bertipe introver. Jumlah akun sebanyak 46 akun ini mendekati jumlah akun yang digunakan oleh [9] sebanyak 50 akun. Total *tweet* yang diperoleh sebanyak 125.897 *tweet*, dengan rata-rata 2.736 *tweet* per akun. Selanjutnya, data profil dan *tweet* dipraproses dan diperoleh nilai (bobot) dari setiap cirinya.

Klasifikasi dilakukan pada perbandingan jumlah data latih dan uji sebesar 70:30, 80:20, 90:10. Pembagian data dilakukan secara acak dengan jumlah data per label kelas diupayakan berimbang. Selanjutnya, dilakukan normalisasi nilai ciri ke dalam rentang -1 s.d. +1. Rentang nilai yang diperoleh dari normalisasi data latih kemudian digunakan untuk menormalisasi data uji.

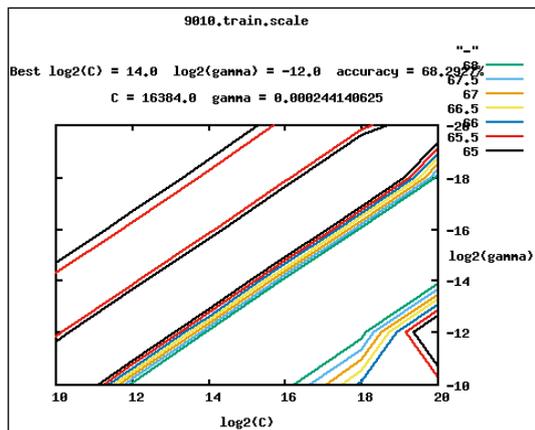
Setelah melakukan *grid search* pada kemungkinan parameter C = $[2^{-5}, \dots, 2^{15}]$ and $\gamma = [2^{-15}, \dots, 2^3]$ dan 10-fold *cross validation* untuk setiap pasangan parameter, dilakukan penyesuaian rentang parameter menjadi C = $[2^{10}, \dots, 2^{20}]$ and $\gamma = [2^{-20}, \dots, 2^{-10}]$. Pasangan parameter terbaik adalah pasangan parameter C=1048576 dan $\gamma=3.814697265625e-06$ dengan akurasi sebesar 83,78% pada perbandingan jumlah data latih dan uji 80:20. Hasil pencarian parameter terbaik diperlihatkan pada Gambar 2, 3 dan 4.



Gambar 2. Hasil pencarian pasangan parameter pada perbandingan 70:30



Gambar 3. Hasil pencarian pasangan parameter pada perbandingan 80:20



Gambar 4. Hasil pencarian pasangan parameter pada perbandingan 90:10

Pasangan parameter terbaik digunakan untuk melatih seluruh data latih dan menghasilkan model. Pada perbandingan 80:20, digunakan data latih sebanyak 37 akun, dan data uji sebanyak 9 akun. Berdasarkan model tersebut, diketahui *support vector* untuk kelas introver sebanyak 8 akun dan kelas

ekstrover sebanyak 7 akun. Pengujian dilakukan terhadap data uji dengan menggunakan model tersebut dan menghasilkan akurasi sebesar 88,89%, dimana terdapat 1 akun dengan label kelas ekstrover yang salah prediksi menjadi kelas introver. Hasil yang diperoleh lebih baik dibandingkan dengan hasil penelitian sebelumnya [18] yang menggunakan metode Naive Bayes Classifier dengan akurasi 83,33%.

Kesimpulan

Penelitian ini berhasil menggunakan metode SVM untuk mengklasifikasikan kepribadian pengguna Twitter ke dalam tipe kepribadian ekstrover atau introver. Fitur yang digunakan, sebanyak 17 fitur, merupakan fitur-fitur yang disesuaikan dengan karakteristik media sosial Twitter. Setelah melakukan *grid search* dan *cross validation*, diperoleh pasangan parameter terbaik yaitu $C=1048576$ dan $\gamma=3.814697265625e-06$ pada pembagian data 80:20. Hasil pengujian dengan model terbaik tersebut memperoleh akurasi sebesar 88,89%.

Saran

Kedepannya, upaya meningkatkan akurasi dapat dilakukan dengan melakukan beberapa hal. Penelitian ini dapat dilanjutkan dengan menggunakan fitur-fitur lain yang belum digunakan sebagaimana tercantum pada [19], dan [16]. Fitur-fitur tersebut memerlukan data tambahan dari akun lain yang menyebutkan (*mention*), *re-tweet*, menyukai status (*favorite*) dari si pemilik akun. Saat ini, fitur-fitur yang digunakan hanya bersumber dari akun itu sendiri. Fitur lain yang dapat digunakan adalah kata-kata yang terdapat di dalam *tweet*. Dengan semakin banyaknya jumlah fitur yang akan digunakan, disarankan untuk menerapkan langkah seleksi ciri dengan menggunakan F-score. Selain itu, jumlah data yang digunakan disarankan untuk diperbanyak, misalnya mencapai 10.000 akun sebagaimana yang digunakan oleh [13]. Hal ini tentunya akan memerlukan sumber daya yang banyak untuk memperoleh *corpus* yang diinginkan tersebut.

Daftar Pustaka

- [1] Jung, C. G., *Psychological Types*, Zurich: Rascher Verlag, 1921.
- [2] Weber, P., *Communication Toolkit for Introverts*: Packt Publishing, 2014.
- [3] Schaubhut, N., Weber, A., dan Thompson, R., *Myers-Briggs Type and Social Media Report*, CPP Inc, 2012.
- [4] Harbaugh, E. R., "The Effect of Personality Styles (Level of

- Introversion- Extroversion) on Social Media Use,” 2010.
- [5] Widiyanti, K. S., dan Herdiyanto, Y. K., “Perbedaan Intensitas Komunikasi Melalui Jejaring Sosial antara Tipe Kepribadian Ekstrovert dan Introvert pada Remaja,” 2013.
- [6] Zywicka, J., dan Danowski, J., “The Faces of Facebookers: Investigating Social Enhancement and Social Compensation Hypotheses; Predicting Facebook and Offline Popularity from Sociability and Self Esteem, and Mapping the Meanings of Popularity with Semantic Networks,” *Journal of Computer-Mediated Communication*, vol. 14, 2008.
- [7] Amiel, T., dan Sargent, S. L., “Individual differences in Internet usage motives,” 2004.
- [8] Golbeck, J., Robles, C., dan Turner, K., “Predicting Personality with Social Media.” pp. 253-262.
- [9] Golbeck, J., Robles, C., Edmondson, M., dan Turner, K., “Predicting Personality from Twitter,” 2011.
- [10] Qiu, L., Han, L., Ramsay, J., dan Yang, F., “You are what you tweet: Personality expression and perception on Twitter,” *Journal of Research in Personality* vol. 46, pp. 710-718, 2012.
- [11] Sarwani, M. Z., dan Mahmudy, W. F., “Analisis Twitter Untuk Mengetahui Karakter Seseorang Menggunakan Algoritma Naïve Bayess Classifier,” 2015.
- [12] Ahmad, N., dan Siddique, J., “Personality Assessment using Twitter Tweets,” 2017.
- [13] Nguyen, T., Phung, D., Adams, B., dan Venkatesh, S., “Towards Discovery of Influence and Personality Traits Through Social Link Prediction,” in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 2011.
- [14] Cortes, C., dan Vapnik, V., “Support-Vector Networks,” 1995.
- [15] Riquelme, F., dan Gonzalez-Cantergiani, P., “Measuring User Influence on Twitter - a Survey,” *Journal of Information Processing and Management*, April 2016, 2016.
- [16] Pal, A., dan Counts, S., “Identifying Topical Authorities in Microblogs,” 2011.
- [17] Samghabadi, N. S., Maharjan, S., Sprague, A., Diaz-Sprague, R., dan Solorio, T., “Detecting Nastiness in Social Media,” in Proceedings of the First Workshop on Abusive Language Online Vancouver, Canada, 2017, pp. 63-72.
- [18] Hamdila, M., “Klasifikasi Kepribadian Ekstrovert dan Introvert pada Akun Twitter Menggunakan Metode Naïve Bayes Classifier,” Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, UIN Suska Riau, Pekanbaru, 2018.
- [19] Chang, C.-C., dan Lin, C.-J., “LIBSVM: A Library for Support Vector Machines,” 2013.