

# ANALISIS TWITTER UNTUK MENGETAHUI KARAKTER SESEORANG MENGGUNAKAN ALGORITMA NAÏVE BAYESS CLASSIFIER

Mohammad Zoqi Sarwani<sup>1)</sup>, Wayan Firdaus Mahmudy<sup>2)</sup>

<sup>1</sup>Magister Ilmu Komputer/Informatika, Program Teknologi Informasi dan Ilmu Komputer, Universitas  
Brawijaya  
Jl.Veteran No.8, Malang, 65145  
Telp : 081515600695, Fax : (0341) 577911  
E-mail : zzaki\_swr@yahoo.com

---

## Abstrak

*Analisis kepribadian seseorang pada beberapa perusahaan maupun instansi merupakan hal yang sangat penting. Analisis tersebut dapat dijadikan pertimbangan dalam perekrutan karyawan maupun untuk kenaikan jabatan. Analisis kepribadian secara konvensional membutuhkan beberapa sumber daya seperti ruangan dan waktu yang cukup lama. Penelitian ini memberikan solusi dengan cukup menggunakan media twitter berdasarkan hasil tweets dari seseorang. Proses klasifikasi menggunakan algoritma Naïve Bayess Classifier. Peneliti menggunakan 10 pengguna twitter sebagai data latih dan 10 pengguna twitter sebagai data uji. Hasil yang diperoleh tersebut kemudian dibandingkan dengan data dari pakar, sehingga didapat keakurasian data mencapai 100%.*

**Kata kunci:** Naïve Bayess Classifier (NBC), twitter, analisis kepribadian

## Abstract

*Trait personality analysis of person on several companies and institutions is very important. That analysis can be considered in the recruitment of employees as well as for promotion. Personality analysis conventionally needs some resources such as rooms and a long time. This study provides a solution by using twitter media based on tweets results of a person. The classification process use Naïve Bayess Classifier algorithm. Researchers used 10 twitter users as training data and 10 users of Twitter as test data. The results of this process compared with data from expert, so that the accuracy of data obtained up to 100%.*

**Keywords:** Naïve Bayess Classifier (NBC), twitter, personality analysis

## 1. PENDAHULUAN

Tes psikologi menjadi salah satu tes yang digunakan sebagai pertimbangan dalam merekrut karyawan atau kenaikan jabatan pada instansi maupun perusahaan. Pengetahuan tentang kepribadian seseorang dianggap penting karena kepribadian tersebut mempengaruhi tingkah laku dalam mengambil keputusan yang berdampak baik atau buruk. Selain itu, kepribadian seseorang dapat dijadikan sebagai salah satu faktor yang digunakan untuk memberi sebuah tanggung jawab pekerjaan.

Salah satu layanan sosial media yang paling populer saat ini adalah *twitter*. *Twitter* telah menghasilkan 110 juta *tweet* setiap hari dan memiliki lebih dari 200 juta pengguna[1]. Banyak penelitian tentang *text mining* menjadikan media sosial sebagai media untuk mendapatkan informasi sentimen ataupun *polling*. Selain itu, *twitter* sering digunakan oleh penggunanya sebagai media untuk mempublikasikan kegiatan sehari-hari atau tempat untuk mencurahkan apa yang dirasakan oleh pengguna. Banyak pengguna *twitter* yang secara tidak sadar memberikan informasi tentang kepribadiannya melalui *tweets* atau posting yang mereka buat dengan bahasa yang alami [2].

Salah satu cara untuk dapat mengetahui kepribadian seseorang yaitu dengan melakukan tes psikologi. Tes psikologi yang dilakukan saat ini kebanyakan melalui tes tulis ataupun tes wawancara yang memakan banyak waktu serta tempat. Saat ini banyak penelitian tentang kepribadian seseorang dilakukan menggunakan media sosial seperti yang dilakukan oleh Barker,dkk[3]. Tulisan tersebut menganjurkan cara untuk memahami apa yang diinginkan orang lain adalah dengan kita mempelajari perilaku alami sehari-hari. Ada banyak tes psikologi dan salah satu yang dianggap akurat adalah tes psikologi berdasarkan indikator MBTI[4]. Pada tes psikologi berdasarkan MBTI terdapat 16 indikator yang kemudian dapat dikelompokkan menjadi 4 kelompok kepribadian[5].

Penelitian sebelumnya dilakukan oleh Qiu,dkk[2] mengenai analisis untuk mengetahui potensi kepribadian seseorang dengan menggunakan *twitter*. Penelitian lain yang melakukan analisis *twitter* untuk sentimen merek produk juga dilakukan oleh Mustofa[6].

Penelitian ini menggunakan *twitter* sebagai media untuk menganalisis kepribadian seseorang. Penelitian sebelumnya telah menunjukkan potensi *twitter* untuk dilakukan penelitian tentang kepribadian seseorang [2]. Dalam proses menganalisis kepribadian seseorang melalui *twitter* dibutuhkan metodologi yang tepat untuk mendapatkan hasil yang akurat. *Tweet* atau posting pada *twitter* merupakan kumpulan kata yang tidak baku sehingga dibutuhkan perlakuan khusus untuk mendapatkan data yang bisa diproses. Oleh sebab itu, dalam pengolahan data diperlukan proses *pre-processing* terlebih dahulu yang kemudian bisa dilakukan pengklasifikasian. Dalam penelitian ini, metode klasifikasi yang dipilih adalah metode klasifikasi *Naïve Bayes Classifier*. Metode tersebut dipilih karena sederhana dan memberikan kemudahan dalam proses pengolahan data serta memberikan tingkat akurasi yang baik. Meskipun sederhana, *Naïve Bayes Classifier* terbukti cukup akurat pada permasalahan klasifikasi berbasis teks seperti penelitian yang telah dilakukan oleh Widodo,dkk [7].

## 2. PENELITIAN TERKAIT

Penelitian tentang *twitter* pernah dilakukan oleh Mustofa[6] untuk menganalisis sentimen suatu *brand*. Penelitian tersebut dilakukan dengan mengambil sampel pengguna *twitter* secara acak. Penelitian lain yang dilakukan oleh Rodiansyah[8] juga menjadikan *twitter* sebagai media penelitiannya.

Penelitian tersebut menggunakan *twitter* untuk menganalisa kemacetan kota Bandung. Metode yang digunakan dalam penelitian [8] adalah metode *Naïve Bayes Classification*.

Penelitian lain yang juga memanfaatkan *twitter* sebagai media dilakukan oleh Qiu,dkk[2]. Penelitian tersebut menganalisis *twitter* untuk mengetahui apakah di dalam *twitter* terdapat karakter kepribadian pengguna *twitter*. Hasil yang diperoleh dalam penelitian tersebut menyatakan bahwa karakter kepribadian seseorang bisa diketahui dari *twitter* mereka.

## 3. METODE

Penelitian ini menggunakan beberapa metode *text mining* dan *data mining* yang telah di rangkum dalam sebuah metodologi. Adapun tahapan yang akan dilakukan dalam penelitian ini adalah sebagai berikut :

### 3.1 Pengumpulan Data

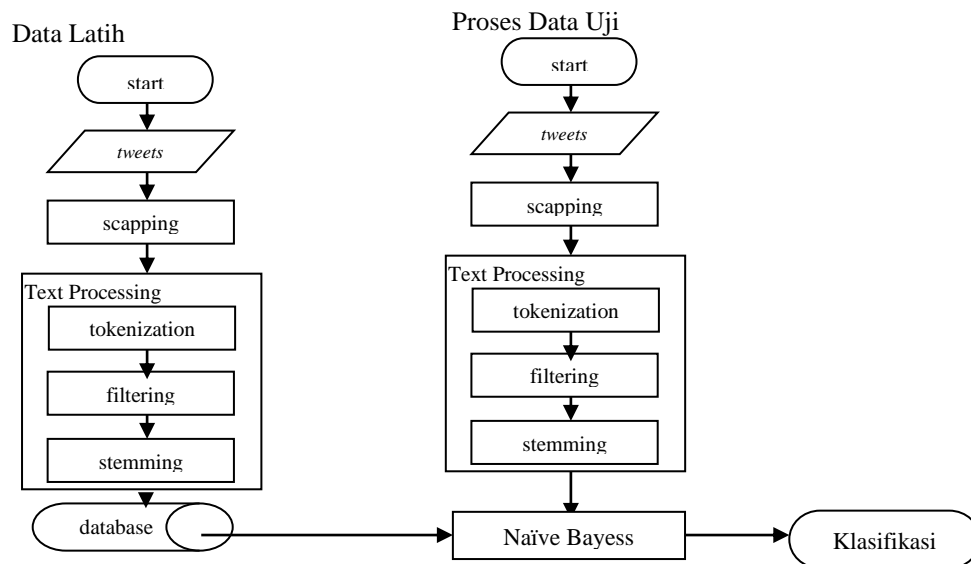
Data yang digunakan pada penelitian ini adalah berupa *tweets* atau posting yang ada pada *twitter*. Teknik pengambilan datanya adalah dengan cara *scrapping* pada *twitter* berdasarkan *user\_id twitter*. Teknik tersebut digunakan untuk mengambil *data tweets* atau postingan pengguna *twitter* yang akan dijadikan sebagai data latih maupun data uji.

### 3.2 Text Mining

*Text mining* merupakan sebuah metode yang digunakan untuk melakukan pemrosesan teks. Pada *text mining* terdapat beberapa tahapan yang dilakukan yaitu *text processing* dan *feature selection*.

#### a. Text processing

*Text Processing* menggunakan beberapa tahapan yang harus dilakukan. Pertama, melakukan *ToLowerCase* yaitu mengubah seluruh data teks menjadi huruf kecil. Setelah itu melakukan *Tokenization* yaitu memecah data yang berupa kalimat menggunakan *delimiter* spasi serta menghilangkan angka dan tanda baca lainnya yang tidak berarti apa-apa pada suatu kata[9].



Gambar 1. Alur kerja metode yang dilakukan

#### b. Future Selection

Pada tahap ini dilakukan proses menghilangkan kata yang dianggap *stopword*. Kemudian dilakukan proses *stemming* untuk menghilangkan kata-kata yang memiliki imbuhan[10][11]. *Stopword* adalah kata yang tidak memiliki arti atau makna yang mencirikan suatu *document*[12], contohnya adalah kata “di”, “dan”, “kamu”, “saya”, “yang”, “oleh” dan seterusnya. *Stemming* adalah proses yang dilakukan untuk melakukan pemetaan dan penguraian bentuk kata sehingga memiliki kata dasar[11].

### 3.3 Naïve Bayes Classification

*Naïve Bayes Classifier* adalah salah satu algoritma yang digunakan untuk klasifikasi teks serta merupakan metode *Machine Learning* yang menggunakan perhitungan probabilitas dan statistik yang dikemukakan oleh Thomas Bayes. Algoritma tersebut digunakan untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa lalu.

Dasar dari *naïve bayes* yang dipakai adalah rumus :

$$P(A|B) = (P(B|A) * P(A))/P(B) \quad (1)$$

Pada pengaplikasiannya rumus ini berubah menjadi :

$$P(C_i|D) = (P(D|C_i) * P(C_i))/P(D) \quad (2)$$

*Naïve Bayes Classifier* adalah model penyederhanaan dari metode *Bayes* yang cocok untuk pengklasifikasian teks atau dokumen. Adapun rumusnya dipaparkan pada Persamaan (3):

$$V_{MAP} = \arg \max P(V_j|a_1, a_2, \dots, a_n) \quad (3)$$

Berdasarkan Persamaan (3) dan Persamaan (1) dapat ditulis :

$$V_{MAP} = \frac{\arg \max_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j) P(v_j)}{P(a_1, a_2, \dots, a_n)} \quad (4)$$

Karena  $P(a_1, a_2, \dots, a_n)$  konstan, Persamaa (1) dapat ditulis menjadi :

$$V_{MAP} = \arg \max_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j) P(v_j) \quad (5)$$

Karena  $P(a_1, a_2, \dots, a_n | v_j) P(v_j)$  sulit dihitung, diasumsikan setiap kata tidak terkait, maka persamaan tersebut ditulis:

$$V_{MAP} = \arg \max_{v_j \in V} P(v_j) \prod_i P(a_i | v_j) \quad (6)$$

$$P(v_j) = \frac{|doc_j|}{|Contoh|} \quad (7)$$

$$P(w_k | v_j) = \frac{n_k + 1}{n + |Konstanta|} \quad (8)$$

Dimana :

$P(v_j)$  : probabilitas setiap dokumen terhadap sekumpulan dokumen

$P(w_k|v_j)$  : Probabilitas kemunculan kata  $w_k$  pada suatu dokumen dengan kategori klas  $v_j$ .  
 $|docs|$  : frekuensi dokumen pada setiap kategori  
 $|Contoh|$  : jumlah dokumen yang ada  
 $N_k$  : Frekuensi kata ke- $K$  pada setiap kategori.  
 Kosakata : jumlah kata pada dokumen tes.

#### 4. STUDI KASUS

Proses perhitungan *naïve bayes classifier* dilakukan beberapa tahap. Pada Tabel 1 menjelaskan tahap perhitungan *naïve bayes* pada perhitungan jumlah setiap kata yang ada pada tweets user tes yang menggunakan rumus Pada Persamaan (5).

Tabel 1 Perhitungan dengan menggunakan rumus pada Persamaan (5)

Kata	Frek Klas Guardian	Frek Klas Artisan	Frek Klas Idealist	Frek Klas Rasional
Twit	0	0	2	0
Nyoba	0	0	1	1
ngetweet	0	1	1	1
Telah	2	2	1	6
terurustiba	0	0	1	0
Piker	1	2	3	2
ngaktifin	0	0	1	0
Alam	1	0	1	1
Share	1	0	1	1
Sini	1	1	1	3
wiuchhh	0	0	1	0
Cnya	0	0	1	0
Hope	0	0	1	0
This	6	0	1	5
Day	3	4	1	1
Will	1	0	1	3
Great	0	1	1	0
Want	1	0	1	1
Free	0	1	1	1
windows	0	0	2	0
Seven	0	0	1	0
Ultimate	0	0	1	0
Join	1	0	1	3
netindonesia	0	0	1	0

Proses selanjutnya adalah menghitung nilai *prior* dengan menggunakan rumus pada Persamaan (8) dan juga menghitung nilai probabilitas pada setiap klas. Adapun hasil dari perhitungan manual menggunakan algoritma *Naïve Bayeslan classifier* dijelaskan pada Tabel 2.

Tabel 2 Hasil perhitungan menggunakan *Naïve Bayess Classifier*

Kata	Nilai Prior untuk Klas Guardian	Nilai Prior untuk Klas Artisan	Nilai Prior untuk Klas Idealis	Nilai Prior untuk Klas Rasional
Twit	0.000603136	0.000658762	0.00233463	0.000343289
Nyoba	0.000603136	0.000658762	0.00155642	0.000686577
ngetweet	0.000603136	0.001317523	0.00155642	0.000686577
Telah	0.001809409	0.001976285	0.00155642	0.002403021
terurustiba	0.000603136	0.000658762	0.00155642	0.000343289
Piker	0.001206273	0.001976285	0.00311284	0.001029866
ngaktifin	0.000603136	0.000658762	0.00155642	0.000343289
Alam	0.001206273	0.000658762	0.00155642	0.000686577
Share	0.001206273	0.000658762	0.00155642	0.000686577
Sini	0.001206273	0.001317523	0.00155642	0.001373155
wiuchhh	0.000603136	0.000658762	0.00155642	0.000343289
Cnya	0.000603136	0.000658762	0.00155642	0.000343289
Hope	0.000603136	0.000658762	0.00155642	0.000343289
This	0.004221954	0.000658762	0.00155642	0.002059732
Day	0.002412545	0.003293808	0.00155642	0.000686577
Will	0.001206273	0.000658762	0.00155642	0.001373155
Great	0.000603136	0.001317523	0.00155642	0.000343289
Want	0.001206273	0.000658762	0.00155642	0.000686577
Free	0.000603136	0.001317523	0.00155642	0.000686577
windows	0.000603136	0.000658762	0.00233463	0.000343289
Seven	0.000603136	0.000658762	0.00155642	0.000343289
Ultimate	0.000603136	0.000658762	0.00155642	0.000343289
Join	0.001206273	0.000658762	0.00155642	0.001373155
Netindonesia	0.000603136	0.000658762	0.00155642	0.000343289
<b>Total Klas Prior</b>	5.77379E-74	3.21225E-74	1.83763E-67	7.40609E-78
<b>Perhitungan Akhir</b>	1.15476E-74	6.42451E-75	5.51288E-68	2.22183E-78
<b>Nilai Probabilitas</b>	2.09465E-07	1.16536E-07	0.999999674	4.03024E-11

Dari data pada Tabel 2 dapat ditarik kesimpulan bahwa kecenderungan karakter pada data tes adalah tergolong dalam klasifikasi karakter Idealis dengan nilai probabilitas 0.999997983.

## 5. HASIL DAN ANALISIS

Pada uji coba yang dilakukan dalam penelitian ini menggunakan 10 pengguna *twitter* sebagai data latih dan 10 pengguna *twitter* sebagai data testing. Dengan menggunakan Persamaan (6) didapatkan hasil yang disajikan pada Tabel 3.

Tabel 3 Data uji hasil klasifikasi dari pakar dan menggunakan *Naïve Bayess Classifier*

Nama Pengguna Twitter	Klasifikasi Dari Pakar	Klasifikasi Menggunakan Naïve Bayess Classiifier
Mahasiswa 1	Guardian	Guardian
Mahasiswa 2	Guardian	Guardian
Mahasiswa 3	Artisan	Artisan
Mahasiswa 4	Idealis	Idealis
Mahasiswa 5	Rasional	Rasional
Mahasiswa 6	Artisan	Artisan
Mahasiswa 7	Idealis	Idealis
Mahasiswa 8	Idealis	Idealis
Mahasiswa 9	Rasional	Rasional
Mahasiswa 10	Artisan	Artisan

Dari Tabel 3 didapatkan hasil klasifikasi karakter pengguna *twitter*. Untuk mengukur tingkat akurasi data hasil klasifikasi menggunakan metode *naïve bayes classifier* dengan cara membagi jumlah data klasifikasi yang benar dengan jumlah seluruh data sehingga diperoleh tingkat akurasi sebesar 100%.

## 6. KESIMPULAN DAN SARAN

Dalam penelitian ini dilakukan analisis terhadap *tweets* atau posting yang terdapat pada *twitter* untuk mengetahui karakter kepribadian seseorang. Hasil penelitian membuktikan bahwa *twitter* dapat digunakan sebagai salah satu media untuk mengetahui kepribadian seseorang melalui posting atau *tweets* mereka. Selain itu, proses pengklasifikasian *twitter* menggunakan metode *Naïve Bayes Classification* juga mampu memberikan tingkat akurasi yang baik dengan membandingkan hasil klasifikasi dari sistem dengan hasil dari pakar.

Untuk pengembangan dari penelitian ini, perlu untuk melakukan percobaan dengan menggunakan jumlah data latih dan data uji yang besar untuk menghitung keakurasian metode yang digunakan dalam penelitian ini.

## 7. DAFTAR RUJUKAN

- [1] Chiang, O. (2011, January 19). Twitter hits nearly 200M accounts, 110M tweets day, focuses on global expansion. *Forbes*.<http://www.forbes.com/sites/oliverchiang/2011/01/19/twitter-hits-nearly-200m-users-110m-tweets-perday-focuses-on-global-expansion/>. *Psychological Science*, 21, 372–374.
- [2] Qiu, L., Lin, H., Ramsay, J., dan Yang, F., 2012. You are what you tweet: Personality expression and perception on Twitter. *Division of Psychology, Singapore. Science Direct*. Pp.710-718.
- [3] Barker, R. G., & Wright, H. S. (1951). *One boy's day: A specimen record of behavior*. Oxford, England: Harper. Borkenau, P., & Liebler, A. (1992). *Trait inferences: Sources of validity at zero*
- [4] Tes Kepribadian. Psikologizone. 2015. 4 Juni 2015.<http://www.psikologizone.com/tes-kepribadian-mbti>
- [5] Keirsey Temperament Sorter. Keirsey.com. 4 Juni 2015.<http://www.keirsey.com/difference.aspx>
- [6] Mustofa, M.M., 2013. More than words: Social network's text mining for consumer brand sentiments. *Science Direct*. Pp.4241–4251.
- [7] Widodo, AW, Mahmudy, WF & Maisuroh, M 2007, 'Klasifikasi artikel otomatis, sebuah kajian eksperimen', *Jurnal Forum Komunikasi Perpustakaan Perguruan Tinggi Negeri (FKP2T)*, vol. 2, no. 1, pp. 39-48.
- [8] Rodiyansyah, S.F., Winarko, Edi, "Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification," vol. 6, no. 1, pp. 91–100, 2012.
- [9] Weiss, S.M., Indurkha, N., Zhang, T., Damerau, F.J. 2005. *Text Mining : Predictive Methods for Analyzing Unstructured Information*. Springer : New York.
- [10] Feldman, R & Sanger, J. 2007. *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press : New York.
- [11] Berry, M.W. & Kogan, J. 2010. *Text Mining Application and theory*. WILEY : United Kingdom.
- [12] Dragut, E., Fang, F., Sistla, P., Yu, S. & Meng, W. 2009. Stop Word and Related Problems in Web Interface Integration.<http://www.vldb.org/pvldb/2/vldb09-384.pdf>. Diakses tanggal 20 juni 2015
- [13] Robbins, Stephen P.; Judge, Timothy A. (2008). *Perilaku Organisasi Buku 1*, Jakarta: Salemba Empat. Hal.126-127