

Analisis Kepribadian Melalui *Twitter* Menggunakan Metode *Logistic Regression* dengan Pembobotan TF-IDF dan AHP

Kartika Prameswari¹, Erwin Budi Setiawan²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹kartikaprameswari@students.telkomuniversity.ac.id, ²erwinbudisetiawan@telkomuniversity.ac.id,

Abstrak

Melakukan analisis kepribadian saat ini menjadi hal yang telah banyak dilakukan. Banyak sekali dalam seleksi pemilihan minat dan karir, klasifikasi kepribadian diikutsertakan. Hal ini bertujuan agar klasifikasi kepribadian sesuai dengan minat dan karir. Namun dalam implementasi untuk melakukan analisis kepribadian memakan waktu yang lama dan memakan biaya yang besar. Tujuan dari penelitian ini adalah untuk memprediksi klasifikasi kepribadian menggunakan cara yang baru yaitu melalui media sosial *twitter*. Analisis ini dilakukan dengan pendekatan linguistik dimana analisis kepribadian melalui bahasa atau kata pada *tweets* yang telah diunggah pengguna *twitter*. Selain menggunakan *tweets*, fitur lain pada *twitter* juga diikutsertakan sebagai bahan pertimbangan untuk penilaian klasifikasi kepribadian melalui pendekatan perilaku sosial. Metode klasifikasi kepribadian ini menggunakan metode *logistic regression* dengan pembobotan TF-IDF dan AHP. Pembobotan TF-IDF dan AHP ini dilakukan untuk menentukan bobot tiap-tiap fitur yang ada pada akun pengguna. Pada pendekatan linguistik digunakan pembobotan TF-IDF, sedangkan untuk pendekatan perilaku sosial menggunakan pembobotan AHP. Hasil dari klasifikasi dengan kedua pembobotan tersebut pada pendekatan perilaku sosial memiliki rata-rata akurasi sebesar 24.95%, sedangkan akurasi tertinggi pada pendekatan linguistik dengan pembagian data set 90:10 adalah 33.5%.

Kata kunci : *Logistic Regression*, AHP, TF-IDF, DISC

Abstract

Conducting personality analysis is now a lot of things done. Many in the selection of interest and career selection, personality classification is included. It aims to make the classification of personality according to interests and career. But in the implementation to conduct personality analysis takes a long time and costs a great cost. The purpose of this research is to predict the classification of personality using a new way that is through twitter social media. This analysis is done with a linguistic approach where the personality analysis through the language or word on tweets that Twitter users have uploaded. In addition to using tweets, other features on Twitter are also included as consideration material in personality classification assessment through social behaviour approach. The personality classification method uses the logistic regression method with the TF-IDF and AHP weighted. The TF-IDF and AHP weighting is done to determine the weight of each feature in the user's account. On the linguistic approach used TF-IDF weighted, while for the approach of social behavior using the weighted AHP. The result of the classification with both of the weighing on the social behavior approach has an average accuracy of 24.95%, while the highest accuracy in the linguistic approach with the division of data set 90:10 is 33.5%.

Keywords: *Logistic Regression*, AHP, TF-IDF, DISC

1. Pendahuluan

Kepribadian merupakan salah satu faktor yang memengaruhi perilaku manusia[1]. *American Psychological Association* (APA) mengemukakan bahwa kepribadian mengacu pada perbedaan individu dalam hal pola karakteristik berpikir, merasakan, dan berperilaku[2]. Tes kepribadian adalah sebuah tes yang dilakukan untuk menentukan tipe kepribadian seseorang[3]. Pengklasifikasian tipe kepribadian yang dilakukan dengan tes kepribadian saat ini sering kali masih menggunakan metode lama dan kurang memanfaatkan teknologi yang ada. Padahal pesatnya perkembangan teknologi memberikan manfaat dalam kehidupan manusia. Klasifikasi kepribadian telah banyak diaplikasikan pada berbagai jurnal ilmiah[4]. Penelitian ini menggunakan *twitter* sebagai media untuk menganalisis kepribadian seseorang[5]. *Twitter* telah menghasilkan 110 juta *tweet* setiap hari dan memiliki lebih dari 200 juta pengguna. Banyak pengguna *twitter* yang secara tidak sadar memberikan informasi tentang kepribadiannya melalui *tweets* atau posting yang mereka buat dengan bahasa alami[6].

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[7]. Dalam melakukan klasifikasi pada penelitian kepribadian ini penulis menggunakan teori DISC.

DISC dibuat untuk membantu individu mengeksplorasi kepribadian dan tipe tingkah laku mereka sehingga mereka dapat lebih baik dalam memahami diri mereka sendiri dan orang lain (Bonnstetter & Suiter, 2011 : 32)[8]. Metode DISC merupakan singkatan dari empat faktor kepribadian yaitu *Dominance*, *Influence*, *Steadness*, dan *Compliance*[9]. Pada penelitian tugas akhir ini, penulis melakukan analisis dalam penilaian kepribadian berdasarkan perilaku sosial dan pendekatan linguistik dari kata-kata *tweets* yang diunggah. Metode yang dipilih untuk melakukan analisis ini adalah *logistic regression* dengan pembobotan TF-IDF dan AHP.

2. Studi Terkait

2.1 Kepribadian

Kepribadian adalah sesuatu sikap atau tingkah laku yang dimiliki oleh seseorang dalam melaksanakan suatu kegiatan yang menjadi tanggung jawabnya untuk menentukan suatu tujuan (Susanna, 2014). Kepribadian setiap individu akan membentuk perilaku individu. Kepribadian merupakan gabungan dari ciri-ciri fisik dan mental yang stabil dari setiap individu yang menandakan identitas individu tersebut sehingga berbeda antara setiap individu yang satu dengan individu lainnya (Robbins & Judge, 2013). Kepribadian merupakan trait bawaan sejak lahir, sedangkan karakter merupakan perilaku hasil pembelajaran (Nofijantie, 2014)[10]. Satu cara yang membantu untuk mendapatkan karakter kepribadian seseorang adalah melalui media sosial. Dengan melihat publikasi kegiatan sehari-hari dapat melihat informasi penting dari pengguna sosial media tersebut. Saat ini sudah banyak penelitian tentang klasifikasi untuk mengetahui karakter seseorang melalui media sosial salah satunya penelitian Barker et al (1992)[11].

2.2 Kepribadian DISC

DISC pertama kali ditemukan oleh dan diperkenalkan oleh William Moulton Marston (1983-1947) seorang psikolog Amerika[13]. DISC adalah sebuah alat ukur psikometri yang mengukur gaya kepribadian seseorang mengenai perilaku kerjanya (Nofiar, 2009: 9)[14]. DISC merupakan model perilaku yang membantu setiap manusia memahami “mengapa seseorang melakukan apa yang dilakukannya”. Disamping itu dinamika dimensi *Dominance*, *Influencing*, *Steadiness*, dan *Conscientiousness* pada setiap orang yang berbeda membentuk model DISC pribadi dan menggambarkan perilaku masing masing[12].

1. *Dominance* (D) Orang yang *Dominance* tinggi akan bersifat asertif (tegas) dan langsung. Biasanya mereka sangat independen, ambisius, gagah serta sangat menyukai tantangan dan persaingan. Dalam pemecahan masalahnya, melakukan pendekatan yang aktif dan cepat menyelesaikan masalah. Mereka dipandang orang lain sebagai orang yang berkemauan keras. Oleh karena itu mereka menginginkan segala sesuatu sesuai dengan kemauan mereka.
2. *Influencing* (I) Tipe *Influencing* ini senang berteman. Mereka suka menghibur orang lain dan bersifat sosial. Dalam penyelesaian masalah atau mengerjakan sesuatu, mereka banyak mengandalkan keterampilan sosial. Orang yang bersifat interpersonal ini senang berpartisipasi dalam kelompok dan suka bekerja sama. Keterbukaan sikapnya membuat orang lain memandang dirinya sebagai pribadi yang gampang bergaul dan ramah. Biasanya pribadi seperti ini memiliki banyak teman.
3. *Steadiness* (S) Orang yang bertipe *Steadiness* ini adalah orang yang berkeras hati, gigih, dan sabar. Mereka mendekati dan menjalani kehidupan dengan memanfaatkan standar yang terukur dan stabil. Pada umumnya mereka tidak begitu suka kejutan dan tidak banyak menuntut dan bersifat akomodatif. Mereka sangat ramah dan memperlihatkan kesetiannya kepada disekitarnya. Orang yang bertipe seperti ini jujur dan sabar. Orang lain memandang mereka sebagai orang yang tenang, berhati-hati dan konsisten dalam cara mereka menjalani kehidupan.
4. *Compliance* (C) Orang pada tipe ini sangat tertarik pada presisi (ketelitian dan kecermatan) dan juga dengan akurasi (kecepatan). Mereka menyukai segalanya serba teratur dan jelas. Mereka sangat fokus terhadap fakta. Orang tipe teliti ini sangat menghargai peraturan. Dalam beraktivitas mereka menggunakan sistematis dan aturan aturan agar semuanya terkelola dengan baik. Mengatasi konflik secara tidak langsung. Dihadapan orang lain, mereka dipandang pasif dan selalu mengalah.

2.3 Penggunaan Fitur

Dalam melakukan penelitian ini, peneliti menggunakan pendekatan linguistik dan pendekatan perilaku sosial. Pendekatan linguistik dalam penelitian ini melalui kata-kata yang diunggah oleh pengguna pada *tweet*. Selain dengan penggunaan linguistik dilakukan juga analisis melalui pendekatan perilaku sosial.

2.3.1 Kepribadian Berdasarkan Pendekatan Perilaku Sosial

Perilaku sosial mendefinisikan kepribadian melalui frekuensi penggunaan media sosial dan tingkat keaktifan antar pengguna (Golbeck dkk, 2011). Fitur yang menunjukkan tingkat perilaku sosial pengguna *Twitter* berdasarkan penelitian yang dilakukan Golbeck (2011) adalah sebagai berikut[15].

- a. *Follower* : *Follower* adalah pengguna *Twitter* lain yang mengikuti pengguna yang diacu. *Follower* dapat melihat *tweet* pengguna yang diacu pada halaman muka akun *Twitter* mereka.
- b. *Friend* : *Friend* adalah pengguna yang diacu menjadi *follower* dari pengguna lain.
- c. *Jumlah mention* : *Mention* yang ditandai dengan '@username' menunjukkan tingkat interaksi pengguna *Twitter* dengan pengguna lain.
- d. *Jumlah hashtag* : *Hashtag* menunjukkan keterlibatan pengguna dengan isu/topik yang sedang dibahas. *Hashtag* ditandai dengan karakter '#'.
- e. *Jumlah reply* : *Reply* adalah mention dari pengguna lain kepada pengguna *Twitter* yang diacu.
- f. *Jumlah URL* : URL adalah tautan berupa informasi *website/blog* yang dicantumkan pengguna.
- g. *Jumlah kata dalam tweet* : *Tweet* adalah tulisan yang terdiri dari kumpulan kata dengan panjang maksimal 140 karakter. Jumlah kata dalam *tweet* adalah total kata yang menyusun *tweet* itu.

Selain menggunakan fitur-fitur seperti yang ada diatas, penulis juga menggunakan fitur lain sebagai bahan penilaian untuk menunjukkan seberapa banyak fitur-fitur lain yang digunakan oleh pengguna serta untuk menunjukkan keaktifan pengguna tersebut di *twitter*. Fitur-fitur tersebut sebagai berikut.

- a. *Jumlah huruf kapital* : Jumlah huruf kapital yang digunakan oleh pengguna pada unggahan *tweets*.
- b. *Jumlah emoji* : Jumlah karakter unik yang ada pada *tweets*. Karakter emoji ini adalah karakter yang dapat menggambarkan sebuah ekspresi sesuai dengan keinginan pengguna.
- c. *Jumlah like* : Banyak jumlah apresiasi dari pengguna lain terhadap *tweet* yang telah diunggah oleh pengguna.
- d. *Jumlah tanda baca* : Banyaknya karakter seperti koma(,), titik(.), tanda(!) dan karakter lainnya yang diunggah pada *tweets*.
- e. *Jumlah media url* : Jumlah pengunggahan tautan video, gambar, atau media lainnya oleh pengguna.
- f. *Jumlah retweet* : Jumlah pengunggahan kembali posting *tweets* dari akun pengguna lain.

2.3.2 Kepribadian Berdasarkan Pendekatan Linguistik

Dalam penelitian analisis kepribadian dilakukan melalui pendekatan linguistik. Pendekatan ini melalui penggunaan kata atau bahasa yang diunggah pada *tweet* yang telah dikumpulkan. *Tweet* dari pengguna yang telah dikumpulkan akan dianalisis untuk menentukan korelasi antara kata atau bahasa yang diunggah dengan klasifikasi kepribadian pengguna *twitter*. Pendekatan linguistik ini tergolong baru karena mencari keterkaitan antara bahasa atau kata dengan klasifikasi kepribadian seseorang. Pada fitur linguistik dilakukan dengan memecah *tweet* menjadi satuan kata dengan pendekatan unigram. Setelah kata dipecah menjadi satuan kata, kata-kata tersebut dibobotkan melalui perhitungan TF-IDF.

2.4 Preprocessing

Preprocessing di sini adalah tahapan awal umum sebelum kumpulan dokumen diklusterisasi. Tujuan dari *preprocessing* antara lain untuk menyeragamkan *dataset* agar lebih mudah diolah. Terdapat beberapa proses umum *preprocessing* yaitu *case folding*, *tokenizing*, *stemming*, dan *filtering*[16].

- a. *Case folding* adalah mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter[17].
- b. *Tokenizing / parsing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya [17].
- c. *Filtering* adalah tahap mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist / stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh *stopwords* adalah "yang", "dan", "di", "dari" dan seterusnya [17].
- d. *Stemming* adalah tahap mencari kata dasar dari tiap kata hasil *filtering*. Pada tahap ini dilakukan proses pengembalian berbagai bentukan kata ke dalam suatu representasi yang sama. Tahap ini kebanyakan dipakai

untuk teks berbahasa inggris dan lebih sulit diterapkan pada teks berbahasa Indonesia. Hal ini dikarenakan bahasa Indonesia tidak memiliki rumus bentuk baku yang permanen [17].

2.5 Pembobotan

Pembobotan merupakan teknik pengambilan keputusan pada suatu proses yang melibatkan berbagai faktor secara bersama-sama dengan cara memberi bobot pada masing-masing faktor tersebut. Pembobotan dapat dilakukan secara objektif dengan perhitungan statistik atau secara subyektif dengan menetapkannya berdasarkan pertimbangan tertentu[19]. Pembobotan kata (*term*) bertujuan untuk memberikan bobot pada setiap kata (*term*) yang terdapat pada dokumen teks yang akan diproses[18]. Pembobotan dalam pembuatan sistem analisis kepribadian ini menggunakan metode TF-IDF dan AHP.

2.5.1 TF-IDF (Term Frequency–Inverse Document Frequency)

Pembobotan fitur merupakan sebuah proses pemberian nilai pada setiap fitur berdasarkan relevansi dan pengaruhnya terhadap hasil klasifikasi. Nilai tersebut nantinya dapat digunakan sebagai dasar untuk melakukan seleksi fitur berdasarkan minimum bobot yang telah dihitung dari setiap fitur. Pembobotan dilakukan dengan menggunakan metode TF-IDF. Algoritma TF-IDF pertama kali dicetuskan oleh Salton dan Buckley pada tahun 1988 dan digunakan untuk kepentingan *information retrieval*, yang kemudian turut dimanfaatkan sebagai salah satu algoritma dalam metode *feature weighting* dalam *text mining*. TF-IDF memiliki formula sebagai berikut[20].Ada pada persamaan (1).

$$TF- IDF = TF \times IDF$$

Rumus tersebut dapat dijabarkan menjadi *term frequency* dari fitur *i* pada dokumen *j* dikalikan dengan IDF dari fitur *i* pada dokumen *j*, dimana IDF sendiri merupakan kepanjangan dari *Inverse Document Frequency*[20]. Ada pada persamaan (2) dan (3).

$$TF = \frac{\text{Jumlah kemunculan term pada satu dokumen}}{\text{Jumlah seluruh term dalam dokumen}}$$

$$IDF = \log \frac{\text{Jumlah seluruh dokumen}}{\text{Jumlah dokumen suatu term muncul}}$$

Semakin sering sebuah fitur muncul dalam sebuah teks, maka semakin besar pula bobot yang akan didapat,yang artinya maka akan semakin penting pula fitur tersebut[20].

2.4.2 AHP (Analytical Hierarchy Process)

AHP adalah metode dalam sistem pengambilan keputusan yang menggunakan beberapa variabel dengan proses analisis bertingkat. Analisis dilakukan dengan memberi nilai prioritas dari tiap-tiap variabel, kemudian melakukan perbandingan berpasangan dari variabel-variabel dan alternatif-alternatif yang ada [21]. Dalam menyelesaikan permasalahan dengan AHP ada beberapa prinsip yang harus dipahami, yaitu [22]:

1. *Decomposition* (membuat hierarki)

Sistem yang kompleks bisa dipahami dengan memecahkannya menjadi elemen-elemen yang lebih kecil dan mudah dipahami

2. *Comparative judgment* (penilaian kriteria dan alternatif) Kriteria dan alternatif dilakukan dengan perbandingan berpasangan sehingga dapat diketahui skala kepentingan dari masing-masing kriteria terhadap kriteria lainnya. Tabel 1 merupakan skala perbandingan yang disajikan oleh Saaty[21]. Pada tulisan ini, Tabel 1 ada pada gambar 1.

Intensitas Kepentingan	Definisi	Keterangan
1	<i>Equal Importance</i> (sama penting)	Kedua elemen mempunyai pengaruh yang sama
3	<i>Weak importance of one over</i> (sedikit lebih penting)	Pengalaman dan penilaian sangat memihak satu elemen dibandingkan dengan pasangannya
5	<i>Essential or strong importance</i> (lebih penting)	Elemen yang satu lebih penting daripada yang lainnya
7	<i>Demonstrated importance</i> (sangat penting)	Satu elemen jelas lebih mutlak penting daripada elemen lainnya
9	<i>Extreme importance</i> (mutlak lebih penting)	Satu elemen mutlak penting daripada elemen lainnya
2,4,6,8	<i>Intermediate values between the two adjacent judgements</i>	Nilai-nilai antara dua nilai pertimbangan-pertimbangan yang berdekatan
Resiprokal	Kebalikan	Jika elemen <i>i</i> memiliki salah satu angka diatas ketika dibandingkan elemen <i>j</i> , maka <i>j</i> memiliki kebalikannya ketika dibanding elemen <i>i</i>

Gambar 1. Skala Penilaian Perbandingan Berpasangan [21]

- 3. *Synthesis of priority* (menentukan prioritas)
- 4. *Logical Consistency* (konsistensi logis)

Secara umum pengambilan keputusan dengan metode AHP didasarkan pada langkah-langkah berikut [21]:

1. Mendefinisikan masalah dan menentukan solusi yang diinginkan, lalu menyusun hierarki dari permasalahan yang dihadapi.
2. Menentukan prioritas elemen
 - a. Langkah pertama dalam menentukan prioritas elemen adalah membuat perbandingan pasangan, yaitu membandingkan elemen secara berpasangan sesuai kriteria yang diberikan.
 - b. Matriks perbandingan berpasangan diisi menggunakan bilangan untuk merepresentasikan kepentingan relatif dari suatu elemen terhadap elemen yang lainnya.
3. Sintesis pertimbangan-pertimbangan terhadap perbandingan berpasangan disintesis untuk memperoleh keseluruhan prioritas. Hal-hal yang dilakukan dalam langkah ini adalah:
 - a. Menjumlahkan nilai-nilai dari setiap kolom pada matriks
 - b. Membagi setiap nilai dari kolom dengan total kolom yang bersangkutan untuk memperoleh normalisasi matriks
 - c. Menjumlahkan nilai-nilai dari setiap baris dan membaginya dengan jumlah elemen untuk mendapatkan nilai rata-rata.
4. Mengukur Konsistensi Dalam pembuatan keputusan, penting untuk mengetahui seberapa baik konsistensi yang ada. Hal-hal yang dilakukan dalam langkah ini adalah sebagai berikut:
 - a. Kalikan setiap nilai pada kolom pertama dengan prioritas relatif elemen pertama, nilai pada kolom kedua dengan prioritas relatif elemen kedua dan seterusnya.
 - b. Jumlahkan setiap baris.
 - c. Hasil dari penjumlahan baris dibagi dengan elemen prioritas relatif yang bersangkutan.
 - d. Jumlahkan hasil bagi di atas dengan banyaknya elemen yang ada, hasilnya disebut λ maks.
5. Melakukan penghitungan *Consistency Index* (CI) dengan rumus[21]. Dijelaskan pada persamaan (4).

$$CI = (\lambda_{max} - n) / n$$

Dimana n = banyaknya elemen.

6. Melakukan penghitungan Rasio Konsistensi/*Consistency Ratio* (CR) dengan rumus[21]. Akan dijelaskan pada persamaan (5).

$$CR = CI/IR$$

Dimana CR = *Consistency Ratio*

CI = *Consistency Index*

IR = *Indeks Random Consistency*

7. Memeriksa konsistensi hierarki. Jika nilainya lebih dari 10%, maka penilaian data *judgment* harus diperbaiki. Namun jika Rasio Konsistensi (CI/CR) kurang atau sama dengan 0,1, maka hasil perhitungan bisa dinyatakan benar, dimana nilai RI atau *random index*, dapat dilihat pada Tabel 1.

Tabel 1. Daftar Index Random Konsistensi [25]

Ukuran Matrik	Nilai (IR)
1,2	0.00
Ukuran Matrik	Nilai (IR)
3	0.58
4	0.90
5	1.12
6	1.24
7	1.32
8	1.41
9	1.45
10	1.49
11	1.51
12	1.48
13	1.56
14	1.57
15	1.59

2.5 Logistic Regression

Regresi Logistik adalah salah satu dari banyak metode pembelajaran mesin yang bekerja dengan mengambil input dan mengalikan nilai input dengan nilai bobot [23]. *Classifier* ini mempelajari fitur dari input yang paling berguna untuk membedakan antara berbagai kelas [24].

a. Komputasi Model Regresi Logistik

Regresi Logistik adalah model diskriminatif yang berarti menghitung $P(y | x)$ dengan membedakan antara nilai-nilai yang mungkin berbeda dari kelas y berdasarkan input yang diberikan x . Persamaan untuk ini adalah seperti yang ditunjukkan di bawah ini[24], dijelaskan pada persamaan (6).

$$P(c | x) = \sum_{i=1}^n w_i \cdot f_i$$

Nilai $P(y | x)$ tidak dapat dihitung secara langsung menggunakan rumus sebelumnya karena akan menghasilkan nilai dari $-\infty$ dan ∞ yang berarti tidak akan memberikan output antara nilai 0 dan 1. Untuk menghasilkan nilai output yang berada di antara nilai 0 dan 1, fungsi \exp berikut digunakan[24], dijelaskan pada persamaan (7).

$$P(c | x) = \frac{1}{2} \exp \sum_{i=1}^n w_i \cdot f_i$$

Untuk mengubah faktor normalisasi Z dan menentukan jumlah fitur N adalah sebagai berikut[24]. Dijelaskan pada persamaan (8).

$$\frac{\exp(\sum_{i=1}^n w_i \cdot f_i)}{\sum_c \exp(\sum_{i=1}^n w_i \cdot f_i)}$$

Dalam pemrosesan bahasa umumnya menggunakan fitur-fitur yang bernilai biner. Fitur tidak hanya properti dari observasi x tetapi juga properti untuk observasi x dan kelas keluaran kandidat c .

2.6 Evaluasi Performasi

Setelah proses pengujian ini selesai dilakukan, diperlukan satu tahap lagi untuk menentukan kualitas dari proses yang sudah dilakukan, yaitu evaluasi hasil. Pada tahapan ini, performa dari perhitungan yang sudah dilakukan akan diuji dengan dua parameter, yaitu *accuracy*, *precision*, dan *recall* (Manning et al., 2009), (Feldman & Sanger, 2007) [25].

a. *Precision* (P) adalah seberapa banyak hasil pemrosesan yang relevan terhadap informasi yang ingin dicari atau dengan kata lain *precision* adalah klasifikasi positif yang benar (*true positive*) dan keseluruhan data yang diprediksikan sebagai kelas positif. *Precision* dapat diperoleh dengan penggunaan persamaan (9).

$$P = \frac{TP}{TP + FP}$$

b. *Recall* (R) adalah seberapa banyak dokumen relevan dalam *collection* yang dihasilkan oleh sistem, atau dengan kata lain *recall* adalah jumlah dokumen yang memiliki klasifikasi positif yang benar (*true positive*) dari semua dokumen yang benar-benar positif (termasuk di dalamnya *false negative*). *Recall* dapat diperoleh dengan menggunakan persamaan (10).

$$R = \frac{TP}{TP + FN}$$

c. Sedangkan *Accuracy* (A) adalah jumlah dokumen yang diklasifikasikan dengan benar, baik itu *true positive* maupun *true negative*. *Accuracy* dapat diperoleh dengan menggunakan persamaan (11).

$$A = \frac{TP + TN}{(TP + FP + TN + FN)}$$

Variabel seperti TN, TP, FN, dan FP berasal dari *confusion matrix*. TN adalah singkatan dari *True Negative*, data negatif yang diklasifikasikan sebagai negatif. TP adalah singkatan dari *True Positif*, data positif yang

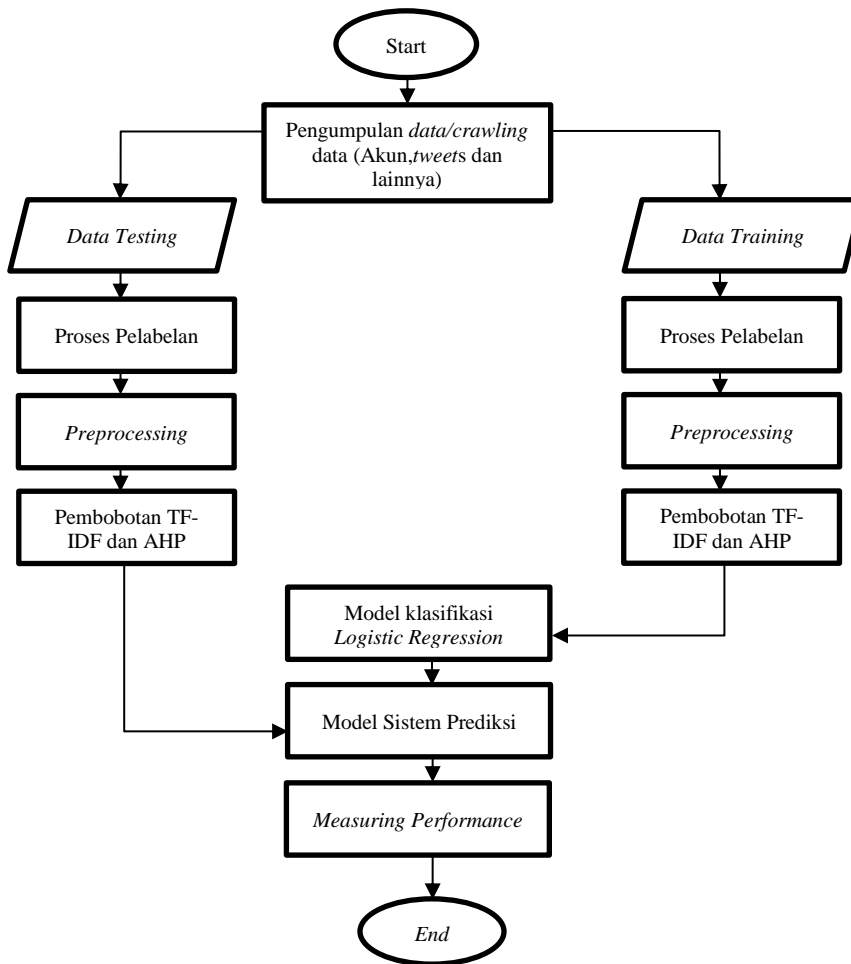
diklasifikasikan sebagai positif. FN adalah singkatan dari *False Negative*, data positif yang diklasifikasikan sebagai negatif. FP adalah singkatan dari *False Positive*, data negatif yang diklasifikasikan sebagai positif [25]. Untuk penjelasan lebih lanjut ada pada Tabel 2.

Tabel 2. Confusion Matrix (Prabowo & Thelwall, 2009)

	Mesin katakan Ya	Mesin katakan Tidak
Manusia katakan Ya	TP	FN
Manusia katakan Tidak	FP	TN

3. Sistem yang Dibangun

Dalam proses pembangunan sistem penulis merancang terlebih dahulu sistem yang akan dibangun. Hal ini dimaksudkan agar sistem berjalan sesuai dengan alur yang sudah dibentuk. Untuk penjelasan mengenai alur sistem prediksi kepribadian yang dirancang dijelaskan pada gambar 2.



Gambar 2.Sistem Prediksi Kepribadian

3.1 Crawling Data

Crawling data pada peneliti ini adalah pengumpulan data pada mahasiswa yang mengikuti tes psikotes Telkom University 2013 melalui media sosial *twitter*. Dari total keseluruhan peserta kurang lebih berjumlah 1600 orang, akun *twitter* yang dapat di-*crawling* berjumlah 454 akun. Data *tweet* yang dikumpulkan sebanyak 1000489 dari keseluruhan akun. Data yang didapatkan dijadikan sebagai data *training* dan data *testing*.

3.2 Pembagian Data

Data yang telah dikumpulkan dibagi menjadi dua. Yaitu data *training* dan data *testing*. Dalam rasio pembagian data dibagi menjadi beberapa rasio. Skenario dalam pembagian data adalah data *training* dan data *testing* (90:10), (80:20), (70:30), (60:40). Pembagian skenario ini digunakan untuk mendapatkan hasil akurasi yang paling tinggi.

3.3 Proses Pelabelan

Proses pelabelan ini berfungsi untuk melabelkan tiap akun sesuai dengan hasil psikotes dari tes. Pelabelan ini diambil dengan mengambil label pertama dari hasil psikotes karena label tersebut yang paling dominan. Sebagai contoh pada pemilik akun bernama Finka Rizkany Tasfin memiliki hasil psikotes CS/DI maka label klasifikasi kepribadian yang dianggap dominan adalah label C (*compliance*).

3.4 Preprocessing

Seluruh data yang telah dikumpulkan sebelum dilakukan analisis terlebih dahulu melalui proses *preprocessing*. Seperti yang telah dijelaskan *preprocessing* ini dilakukan dalam beberapa tahapan yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*. Setelah proses *preprocessing*, data-data akan memasuki tahap pembobotan menggunakan perhitungan TF-IDF.

3.5. Pembobotan TF-IDF dan AHP

Pada proses ini data *tweet* yang telah dikumpulkan akan melalui tahap pembobotan. Tahap pembobotan ini dilakukan untuk mendapatkan bobot dari tiap-tiap kata yang ada di dalam *tweet*. Pada pembobotan *tweet* ini dilakukan menggunakan metode TF-IDF. Untuk fitur perilaku sosial dilakukan melalui metode AHP. Pada metode AHP fitur-fitur yang digunakan sebanyak 15 fitur seperti *followers*, *following*, jumlah *tweet*, *like*, *emoji*, Panjang bio, *website*, media URL, *retweet*, *hashtag*, Panjang *tweet*, *mean* kata dan *mention*. Penggunaan 15 fitur dilakukan agar mendapatkan hasil fitur yang paling berpengaruh.

3.6 Model Logistic Regression

Tahap pada proses ini adalah tahapan yang paling penting karena tahap ini untuk mengklasifikasikan data yang telah melalui proses sebelumnya. Proses ini menggunakan metode *logistic regression* untuk mengklasifikasikan data. Tahap ini akan memprediksi apakah data yang telah diinputkan sesuai atau tidak.

3.7 Evaluasi Performasi

Tahap evaluasi performasi adalah tahapan akhir untuk melihat besar akurasi, *precision*, *recall*. Evaluasi ini dilakukan untuk melihat bagaimana performansi dari sistem yang telah dibuat.

4. Hasil Analisa dan Uji

Pada bagian ini akan dijelaskan hasil pengujian yang telah dibangun sesuai dengan rangkaian sistem yang telah dibuat.

4.1 Data Set

Data dari hasil tes psikotes mahasiswa Telkom University 2013, didapatkan akun pengguna yang berhasil dikumpulkan berjumlah 454 akun dari keseluruhan kurang lebih 1600 peserta. Tidak terkumpulnya akun sesuai dengan jumlah keseluruhan peserta dikarenakan adanya akun peserta yang dinonaktifkan, akun dalam keadaan di privasi dan akun yang nama penggunanya tidak sesuai dengan nama peserta. Contoh hasil dari tes psikotes mahasiswa Telkom University 2013 sebagai berikut.

Tabel 3. Hasil Psikotes Mahasiswa Universitas Telkom 2013

No	NAMA MAHASISWA	DISC
1	ADITIA DITA ARIFIN	SI/CD
2	ASRI WIDIASTUTI	CSD/I

Dari hasil psikotes yang didapatkan diambil label pertama pada hasil untuk mendapatkan sifat kepribadian yang paling dominan. Contohnya, pada mahasiswa yang bernama Aditia Dita Arifin memiliki hasil psikotes SI/CD maka sifat yang paling dominan adalah pada kelas S (*Steadiness*). Pengambilan label pertama ini dikarenakan analisis kepribadian ini hanya menggunakan 4 kelas kategori yaitu, kelas D (*Dominance*), kelas I (*Influence*), kelas S (*Steadiness*) dan kelas C (*Compliance*) untuk menentukan klasifikasi kepribadian. Selanjutnya setelah data pada hasil psikotes dilakukan pengumpulan akun *twitter* peserta psikotes. Proses ini disebut juga dengan *crawling*. Salah satu contoh *crawling* data ada pada Tabel 4.

Tabel 4. Hasil Crawling Akun Twitter Peserta Psikotes Mahasiswa Telkom University 2013

No	Nama	Akun	Follower	Following	Jumlah Tweet	Like	Label
1	Sheyla Putri Nevertari	Sheylaputnev	842	483	24020	81		Dominance
2	Okdi Rachmadian	Heyokday	344	602	10004	92		Influence
3

Setelah dilakukan *crawling* data jumlah mahasiswa yang mengikuti tes psikotes mahasiswa Telkom University 2013 dikumpulkan dan dilabelkan sesuai dengan label DISC. Jumlah peserta sesuai dengan label DISC ada pada Tabel 5.

Tabel 5. Jumlah Peserta Sesuai dengan Label DISC

No	Label	Jumlah
1	D	19
2	I	119
3	S	182
4	C	134

4.2 Hasil dan Analisis

Pada bagian ini, dijelaskan bagaimana hasil uji dari sistem yang telah dibangun sesuai dengan gambaran rancangan sistem yang telah dibuat sebelumnya.

4.2.1 Hasil Preprocessing dan Pembobotan

Pada tahap ini, data *tweet* yang telah dikumpulkan dilakukan proses *preprocessing*. Proses *preprocessing* ini dilakukan data untuk menghilangkan data yang tidak sempurna dan membuat data menjadi lebih terstruktur. Proses ini memiliki empat tahap *case folding*, *tokenizing*, *stemming*, dan *filtering*. Pada tahap *case folding* data pada *tweet* yang memiliki huruf kapital diseragamkan dengan mengganti huruf kapital menjadi huruf kecil. Pada tahap *tokenizing* tiap-tiap karakter tanda baca dan memecah kata menjadi satuan kata dengan pendekatan unigram. Pada tahap *filtering* kata-kata akan dipilih mana kata yang paling penting. Tahap terakhir adalah *stemming* dimana kata yang memiliki kata imbuhan diseragamkan menjadi kata dasar.

Tabel 6. Tabel Hasil Preprocessing

Akun : @ayindaf	Sebelum Preprocessing	Salah siapa itu ya ampun
	Setelah Preprocessing	Salah siapa itu

Setelah *tweet* melalui tahap *preprocessing* tiap-tiap kata yang telah di-*preprocessing* kata-kata diberi bobot menggunakan metode pembobotan TF-IDF. Berikut contoh gambaran pembobotan kata menggunakan metode TF-IDF.

Tabel 7. Tabel Hasil Preprocessing

Akun	Kata							
	Lama	Rasa	Harap	Bisa	Ajar	Ingin	Sikap	Benci
TF 1	34	44	9	145	40	9	0	4
TF 2	20	25	4	35	6	4	1	3
TF 3	1	0	0	0	0	0	0	0
IDF	0.0796	0.0671	0.1081	0.0443	0.0889	0.0983	0.3068	0.2421
TF-IDF 3	0.0796	0	0	0	0	0	0	0

Untuk pembobotan dalam pendekatan linguistik digunakan dengan pembobotan TF-IDF sedangkan untuk pendekatan perilaku sosial digunakan pembobotan perhitungan menggunakan metode AHP. Dalam menentukan bobot hal pertama dilakukan adalah menentukan matriks perbandingan berpasangan. Hal ini dilakukan untuk menentukan bobot prioritas diperlukan untuk melakukan perhitungan selanjutnya yaitu menentukan nilai *Eigen value*. Gambaran perhitungan matriks perbandingan berpasangan dan hasil nilai *Eigen value* ada pada halaman lampiran.

Setelah perhitungan bobot didapatkan kemudian digabungkan dengan data-data yang ada pada akun *twitter* yang telah dikumpulkan untuk dilakukan perhitungan dengan bobot AHP. Untuk hasil dari pembobotan AHP pada tiap akun dilampirkan pada halaman lampiran.

4.2.2 Hasil Akurasi Klasifikasi Logistic Regression

Setelah melalui tahap *preprocessing* kata-kata tersebut masuk ke dalam tahap klasifikasi *logistic regression*. Pada *logistic regression* proses klasifikasi dihitung dengan menghitung peluang dari klasifikasi DISC dari kata-kata yang telah dibobotkan. Dari kata-kata tersebut dibentuk pengujian melalui beberapa skenario. Hasil akurasi dari skenario sebagai berikut.

Tabel 8. Tabel Hasil Akurasi Logistic Regression dan TF-IDF

Hasil Akurasi TF-IDF		
No	Data Set	Akurasi
1	90:10	33.5%
2	80:20	29.7%
3	70:30	29.8%
4	60:40	29.9%

Untuk pendekatan perilaku sosial hasil klasifikasi didapatkan dengan menggunakan *software* SPSS. *Software* ini akan menentukan prediksi pada pembobotan AHP dengan klasifikasi *logistic regression*. Dari hasil tersebut didapatkan hasil akurasi sebagai berikut.

Tabel 9. Tabel Hasil Akurasi Logistic Regression dan AHP

Label	Jumlah Data	Akurasi
D	19	4.2%
I	119	26.2%
S	182	40.1%
C	134	29.5%

Dalam pendekatan perilaku sosial ini dilakukan menggunakan 15 fitur. Fitur-fitur ini dimaksudkan untuk mendapatkan hasil akurasi yang paling baik. Dari perhitungan fitur paling berpengaruh adalah *followers*. Fitur *followers* ini berpengaruh karena memiliki *rank* paling tinggi, dimana nilai dari bobot paling besar.

4.3 Hasil Uji

Hasil Uji dari seluruh percobaan dengan menggunakan fitur, skenario pembagian data *set*. Akurasi terbaik dari pendekatan *linguistic* melalui pembobotan TF-IDF sebesar 33.55% dengan pembagian data *set* data uji dan data latih 90:10. Sedangkan pada hasil pembobotan AHP untuk pendekatan perilaku sosial, didapatkan akurasi terbaik pada label I dengan fitur yang digunakan sebanyak 15 fitur yaitu *followers*, *following*, jumlah *tweet*, *like*, *emoji*, panjang *bio*, *website*, media URL, *retweet*, *hashtag*, panjang *tweet*, *mean* kata, dan *mention* dengan akurasi sebesar 40.1%. Bobot prioritas yang paling berpengaruh dalam pembobotan AHP adalah *followers*.

5. Kesimpulan

Pada penelitian analisis kepribadian melalui twitter dengan *logistic regression* dilakukan dengan menggunakan dua pembobotan. Metode pembobotan yang digunakan yaitu TF-IDF dan AHP. Pembobotan TF-IDF digunakan untuk pendekatan linguistik, dimana pendekatan ini akan menganalisis melalui kata atau bahasa pada *tweet* yang diunggah. Pembagian skenario dibagi menjadi empat yaitu dengan membagi data *training* dan data *testing* (90:10), (80:20), (70:30), (60:40). Pembagian skenario ini digunakan untuk mendapatkan hasil akurasi yang paling tinggi. Pada pembagian rasio data *training* dan data *testing* 90:10 didapatkan hasil dengan akurasi yang paling tinggi yaitu 33.5%. Pada pendekatan perilaku sosial digunakan pembobotan AHP. Dari pembobotan AHP hasil yang didapatkan data dengan akurasi terbesar pada data label S (*Steadiness*) sebesar 40.1%. Dalam melakukan percobaan ini akurasi yang didapatkan jauh dari yang diharapkan. Kesalahan-kesalahan yang menjadi alasan akurasi yang didapatkan tidak sesuai harapan karena besarnya data yang di-*input* membuat data yang tidak terstruktur masuk kedalam sistem analisis. Saran untuk penelitian selanjutnya adalah mencari algoritma yang dapat melakukan penyaringan data untuk jumlah data yang besar.

Daftar Pustaka

- [1] Chaplin JP. Kamus Lengkap Psikologi. Jakarta: Raja Grafindo Persada, 2011.
- [2] American Psychological Association (APA). Personality. [cited 22 Agustus 2016]. Available from: <http://www.apa.org/topics/personality>
- [3] F. Littauer, Personality Plus, Tangerang: Karisma Publishing Group, 2011.
- [4] Yusuf Dwi Santoso, Suhartono. 2017. Sistem Klasifikasi Tipe Kepribadian dan Penerimaan Teman Sebaya Menggunakan Jaringan Syaraf Tiruan Backpropagation. Vol. III No. 1
- [5] 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.
- [6] 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-200musers-110m-tweets-perday-focuses-on-global-expansion/>. Psychological Science, 21, 372–374
- [7] Mohammad Zoqi Sarwani1), Wayan Firdaus Mahmudy. 2015. Analisis Twitter Untuk Mengetahui Karakter Seseorang Menggunakan Algoritma Naive Bayes Classifier. Seminar Nasional Sistem Informasi Indonesia, 2-3 November 2015
- [8] Andes Wardy, Masydzulhak Djamil. 2013. Analisis Profile Kepribadian Menggunakan DISC Yang Mempengaruhi Penyelesaian Studi Mahasiswa Magister Manajemen UMB Kampus Kranggan. Vol 3 No 1.
- [9] Perancangan Sistem Pakar Menggunakan Metode Bayes Untuk Mengenal Kepribadian Dengan Metode DISC
- [10] Sri Sarjana, Nur Khayati. 2016. Pengaruh Etika, Perilaku, Dan Kepribadian Terhadap Integritas Guru. Jurnal Pendidikan dan Kebudayaan, Vol. 1, Nomor 3
- [11] Yessivha Imanuela Claudy, Rizal Setya Perdana, M. Ali Fauzi. 2018. Klasifikasi Dokumen Twitter Untuk Mengetahui Karakter Calon Karyawan Menggunakan Algoritma K-Nearest Neighbor (KNN). Vol.2 No. 8
- [12] Rika Rachmawati, Dyah Kusumastuti, Neneng Susanti. 2016. Pembelajaran Soft Skill Berbasis Pada Hasil Pemetaan Personality Lulusan Universitas Widyatama Dengan Pendekatan Metode DISC (Dominance, Influence, Steadiness Dan Compliance)
- [13] Aline Embun Pramadhani, Tedy Setiadi. (2014). Penerapan Data Mining Untuk Klasifikasi Prediksi Penyakit ISPA (Infeksi Saluran Pernapasan Akut) Dengan Algoritma Decision Tree (ID3). Jurnal Sarjana Teknik Informatika Volume 2 Nomor 1. Page 833
- [14] Septyani Mufida. 2012. Perbedaan Burn out Ditinjau Dari Gaya Kepribadian Dominance, Influence, Steadiness, Dan Compliance
- [15] Agnes Theresia Damanik. 2015. Prediksi Kepribadian Big 5 Pengguna Twitter dengan Support Vector Regression. Vol. 3 No. 1
- [16] Amalia, Maya Silvi Lydia, Siti Dara Fadillah, dan Miftahul Huda. 2018. Perbandingan Metode Kluster dan Preprocessing untuk Dokumen Berbahasa Indonesia. Vol. 14, No. 1
- [17] Diah Pudi Langgeni, ZK. Abdurahman Baizal, Yanuar Firdaus A.W. 2010. Clustering Artikel Berita Berbahasa Indonesia Menggunakan Unsupervised Feature Selection
- [18] Winda Estu Nurjanah, Rizal Setya Perdana, Mochammad Ali Fauzi. (2017) Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 1, No. 12. Page 1751
- [19] Spatial Database Analysis Facilities (SDAF) LABORATORIUM ANALISIS LINGKUNGAN DAN PERMODELAN SPASIAL DEPARTEMEN KONSERVASI SUMBERDAYA HUTAN DAN EKOWISATA FAKULTAS KEHUTANAN INSTITUT PERTANIAN BOGOR. Hal 56
- [20] Garnis Berliana1, Shaufiah, S.T, M.T., Siti Sa'adah, S.T., M.T. 2018 Klasifikasi Posting Tweet mengenai Kebijakan Pemerintah Menggunakan Naive Bayesian Classification. Vol.5, No.1 Maret 2018
- [21] Saaty, T.L. 2008. Decision Making with Analytic Hierarchy Process. International Journal Services Sciences Vol. 1 No 1.
- [22] Saaty, T.L. 2000. The Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process. Pittsburgh: RWS Publication University of Pittsburgh.
- [23] S Raschka "Python Machine Learning", dalam Python Machine Learning Birmingham, Packt Publishing Ltd, 2015.
- [24] D. Jurafsky J Martin "Speech and Language Processing (3rd ed draft)" 24 August 2015 [online] <https://web.stanford.edu/~jurafsky/slp3/7.pdf>.

- [25] Nurrin Muchammad Shiddieqy Hadna , Paulus Insap Santosa , Wing Wahyu Winarno. 2016. STUDI LITERATUR TENTANG PERBANDINGAN METODE UNTUK PROSES ANALISIS SENTIMEN DI TWITTER

Lampiran

Matriks perbandingan berpasangan (pairwise comparison) dalam desimal

	Kriteria	Follower	Following	Jumlah_Tweet	Jumlah_Like	Jumlah_Emoji	Panjang_Bio	Website	Media_URL	Retweet	Hashtag	Panjang_Tweet	Mean_PT	Karakter_Tweet	Mean_Tweet	Mention	Eigen Value	Bobot Prioritas
1	Follower	1.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	4.5425	0.2308
2	Following	0.1429	1.0000	0.2000	0.2000	0.5000	0.5000	3.0000	3.0000	0.2000	5.0000	0.2000	0.5000	0.2000	0.3333	0.2000	0.5485	0.0276
3	Jumlah_Tweet	0.1429	5.0000	1.0000	0.5000	3.0000	3.0000	5.0000	3.0000	3.0000	5.0000	5.0000	0.2000	5.0000	2.0000	0.5000	1.5764	0.0801
4	Jumlah_Like	0.1429	5.0000	2.0000	1.0000	3.0000	5.0000	5.0000	3.0000	3.0000	5.0000	5.0000	5.0000	5.0000	3.0000	0.3333	1.6946	0.0861
5	Jumlah_Emoji	0.1429	2.0000	0.3333	0.3333	1.0000	5.0000	0.5000	0.5000	0.3333	2.0000	3.0000	5.0000	0.5000	0.3333	0.3333	0.8090	0.0234
6	Panjang_Bio	0.1429	2.0000	0.3333	0.2000	0.2000	1.0000	0.5000	0.5000	0.2000	2.0000	0.3333	0.5000	0.2000	0.3333	0.2000	0.4599	0.0234
7	Karakter_Bio	0.1429	0.2000	0.2000	0.2000	0.5000	0.5000	0.5000	0.5000	0.2000	5.0000	0.2000	0.5000	0.2000	0.3333	0.2000	0.3584	0.0182
8	Website	0.1429	0.3333	0.2000	0.2000	2.0000	2.0000	1.0000	0.5000	0.2000	0.5000	0.2000	0.5000	0.3333	0.3333	0.2000	0.4599	0.0234
9	Media_URL	0.1429	0.3333	0.3333	0.3333	2.0000	2.0000	2.0000	1.0000	0.2000	2.0000	0.2000	0.5000	0.3333	0.3333	0.2000	0.5678	0.0289
10	Retweet	0.1429	5.0000	0.3333	0.3333	3.0000	5.0000	5.0000	5.0000	1.0000	5.0000	0.2000	0.5000	0.3333	0.5000	0.2000	0.9618	0.0489
11	Hashtag	0.1429	0.2000	0.2000	0.2000	0.5000	0.5000	2.0000	0.5000	0.2000	1.0000	0.3333	0.5000	0.2000	0.5000	0.2000	0.4233	0.0215
12	Panjang_Tweet	0.1429	5.0000	0.2000	0.2000	3.0000	5.0000	5.0000	3.0000	3.0000	5.0000	1.0000	0.2000	2.0000	0.5000	0.2000	1.1089	0.0563
13	Mean_PT	0.1429	2.0000	0.2000	0.2000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	5.0000	1.0000	0.3333	0.3333	0.5000	0.9153	0.0465
14	Karakter_Tweet	0.1429	5.0000	0.2000	0.2000	2.0000	5.0000	3.0000	3.0000	3.0000	5.0000	0.5000	3.0000	1.0000	0.5000	0.2000	1.1024	0.0560
15	Mean_Tweet	0.1429	3.0000	0.5000	0.3333	3.0000	3.0000	3.0000	2.0000	2.0000	2.0000	3.0000	2.0000	1.0000	1.0000	0.5000	1.3153	0.0668
16	Mention	0.1429	5.0000	2.0000	3.0000	3.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	2.0000	5.0000	2.0000	1.0000	2.3275	0.1183
17	Tanda_Baca	0.1429	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2567	0.0130
18	Huruf_Kapital	0.1429	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2567	0.0130
	Total	3.4286	48.4667	15.6333	14.8333	36.1000	51.9000	49.9000	42.9000	30.9333	54.4000	35.5667	30.3000	30.0333	19.7333	12.3667	19.6801	0.9823

Perhitungan Matrix Perbandingan Berpasangan AHP Menggunakan MS.Excel

Perhitungan Bobot Prioritas Menggunakan MS. Excel

Menentukan Bobot Prioritas

	Kriteria	Follower	Following	Jumlah_Tweet	Jumlah_Like	Jumlah_Emoji	Panjang_Bio	Website	Media_URL	Retweet	Hashtag	Panjang_Tweet	Mean_PT	Karakter_Tweet	Mean_KT	Mention	Jumlah	Prioritas
1	Follower	0.3333	0.1462	0.4656	0.4918	0.1989	0.1373	0.1429	0.1667	0.2308	0.1308	0.2002	0.2381	0.2378	0.3684	0.5949	4.0837	0.2722
2	Following	0.0476	0.0209	0.0133	0.0141	0.0142	0.0098	0.0612	0.0714	0.0066	0.0935	0.0057	0.0170	0.0068	0.0175	0.0170	0.4166	0.0278
3	Jumlah_Tweet	0.0476	0.1045	0.0665	0.0351	0.0852	0.0588	0.1020	0.0714	0.0989	0.0935	0.1430	0.0068	0.1699	0.1053	0.0425	1.2310	0.0821
4	Jumlah_Like	0.0476	0.1045	0.1330	0.0703	0.0852	0.0980	0.1020	0.0714	0.0989	0.0935	0.1430	0.1701	0.1699	0.1579	0.0283	1.5736	0.1049
5	Jumlah_Emoji	0.0476	0.0418	0.0222	0.0234	0.0284	0.0980	0.0102	0.0119	0.0110	0.0374	0.0858	0.1701	0.0170	0.0175	0.0283	0.6506	0.0434
6	Panjang_Bio	0.0476	0.0418	0.0222	0.0141	0.0057	0.0196	0.0102	0.0119	0.0066	0.0374	0.0095	0.0170	0.0068	0.0175	0.0170	0.2849	0.0190
7	Website	0.0476	0.0070	0.0133	0.0141	0.0568	0.0392	0.0204	0.0119	0.0066	0.0093	0.0057	0.0170	0.0113	0.0175	0.0170	0.2948	0.0197
8	Media_URL	0.0476	0.0070	0.0222	0.0234	0.0568	0.0392	0.0408	0.0238	0.0066	0.0374	0.0057	0.0170	0.0113	0.0175	0.0170	0.3734	0.0249
9	Retweet	0.0476	0.1045	0.0222	0.0234	0.0852	0.0980	0.1020	0.1190	0.0330	0.0935	0.0057	0.0170	0.0113	0.0263	0.0170	0.8058	0.0537
10	Hashtag	0.0476	0.0042	0.0133	0.0141	0.0142	0.0098	0.0408	0.0119	0.0066	0.0187	0.0095	0.0170	0.0068	0.0263	0.0170	0.2578	0.0172
11	Panjang_Tweet	0.0476	0.1045	0.0133	0.0141	0.0852	0.0980	0.1020	0.1190	0.0989	0.0935	0.0286	0.0068	0.0680	0.0263	0.0170	0.9228	0.0615
12	Mean_PT	0.0476	0.0418	0.0133	0.0141	0.0568	0.0392	0.0408	0.0476	0.0659	0.0374	0.1430	0.0340	0.0113	0.0175	0.0425	0.6529	0.0435
13	Karakter_Tweet	0.0476	0.1045	0.0133	0.0141	0.0568	0.0980	0.0612	0.0714	0.0989	0.0935	0.0143	0.1020	0.0340	0.0263	0.0170	0.8529	0.0569
14	Mean_KT	0.0476	0.0627	0.0333	0.0234	0.0852	0.0588	0.0612	0.0714	0.0659	0.0374	0.0572	0.1020	0.0680	0.0526	0.0425	0.8693	0.0580
15	Mention	0.0476	0.1045	0.1330	0.2108	0.0852	0.0980	0.1020	0.1190	0.1648	0.0935	0.1430	0.0680	0.1699	0.1053	0.0850	1.7297	0.1153

Perhitungan CR (Consistency Ratio) Kriteria Menggunakan MS. Excel

Perhitungan CR Kriteria

	Kriteria	Jumlah Per Baris	Prioritas	Hasil
1	Follower	5.2363	0.2722	5.5085
2	Following	0.4819	0.0278	0.5097
3	Jumlah_Tweet	1.6430	0.0821	1.7251
4	Jumlah_Like	2.0226	0.1049	2.1276
5	Jumlah_Emoji	0.7669	0.0434	0.8103
6	Panjang_Bio	0.3339	0.0190	0.3529
7	Website	0.3286	0.0197	0.3483
8	Media_URL	0.4114	0.0249	0.4363
9	Retweet	0.8415	0.0537	0.8952
10	Hashtag	0.3269	0.0172	0.3441
11	Panjang_Tweet	1.0549	0.0615	1.1164
12	Mean_PT	0.9061	0.0435	0.9496
13	Karakter_Tweet	0.9567	0.0569	1.0136
14	Mean_KT	1.0931	0.0580	1.1511
15	Mention	2.2780	0.1153	2.3933
				19.6819

Perhitungan *Logistic Regression* dan AHP dengan SPSS 24 Nominal Regression

[DataSet2]

Case Processing Summary

	N	Marginal Percentage
label 0	19	4.2%
1	119	26.2%
3	182	40.1%
4	134	29.5%
Valid	454	100.0%
Missing	0	
Total	454	
Subpopulation	454 ^a	

a. The dependent variable has only one value observed in 454 (100.0%) subpopulations.

Tampilan *Crawling Data Akun*

Dashboard Control panel

Data Crawling Process

Term Weighting TF-IDF / AHP Process

Do Crawling

Account Name
Enter Account

Total Tweet
Enter Total

Simpan data hasil crawling ke dalam database

Crawl

Crawling Result

Akun Followers: 24 Following: 177 Total Tweet: 1032

siojanpaujan "my life my pain :)"

siojanpaujan "Sekian \"pipebomb\" dari gue udah lama juga ane ga ngefrontal dan gue merasa gokil kembali.....bye"

siojanpaujan "Gue harap gue bisa belajar dari pengalaman dan ingin merubah sikap ane benci klo disamakan dg orang yg ga ngerti kehidupan....."

siojanpaujan "Frontal boleh, ini demi kebaikan dan kesadaran kok asal jangan kelewat gblk aja wahahaha....."

siojanpaujan "<Ponytale> PUKULAN 2 DIMENSI #LINEWebtoon|nhttp://t.co/5DkzhwbT6M"

siojanpaujan "Ga pernah bersuara sekalinya bersuara omonganya nyelekit dihati, oh...kepribadian macem apa ini?"

siojanpaujan "http://t.co/U9don6PSX2"

siojanpaujan "Akan ada penyesalan terbesar manusia dan tak bisa mengubahnya"

siojanpaujan "Happy sadnight asand you are the dumbest boy hahaha.... shame you....."bercermin"

siojanpaujan "@RealDhura va kali indomie hikin semuk ane hampir tian hari makan ala hadannya teten sini sini ala. slim :v"

Perhitungan Term Kata

DISCPrediction Kartika Prameswari

Term Weighting Control panel Home - Term Weighting

No	Akun	lama	rasa	harap	bisa	ajar	ingin	sikap	benci	hidup	frontal	baik	sadar	asal	jangan	suara	hati	sesal	besar	tak	ya	gemuk	jadi	bimbang	mau	suka	tarik	nafsu	kurang	sedang	bahas	pikir	akrab	glat	maah
1	TF sijanpaujan	20	25	4	35	6	4	1	3	10	2	14	2	5	14	5	11	1	3	11	35	6	56	1	19	13	5	2	6	7	4	11	3	3	21
2	TF Gandungtaris	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	TF hilmynoki	54	26	6	105	11	12	1	2	12	0	21	6	7	47	9	15	2	11	18	121	1	111	0	105	12	4	1	10	6	6	4	0	1	52
4	TF Juansalautsar	34	44	9	145	40	9	0	4	20	1	23	4	8	55	7	17	1	13	12	92	0	172	1	23	26	5	1	20	11	16	6	0	1	72
5	TF chirstin	44	30	6	126	15	5	0	3	7	0	13	2	6	50	4	12	1	2	19	138	3	120	0	90	34	0	0	5	7	0	8	0	1	53
6	TF Dinindiyani_W	25	59	9	108	10	15	9	2	9	0	45	11	2	66	9	68	3	3	48	200	1	130	0	143	58	1	0	12	2	2	9	2	1	60
7	TF malindoy	23	19	6	116	9	6	0	1	14	0	39	7	1	34	1	34	0	5	16	225	1	66	0	140	42	5	1	2	4	0	3	0	1	17
8	TF mogomo21	29	50	2	150	29	2	1	0	12	1	12	3	1	40	5	10	0	8	7	169	0	180	0	142	22	6	1	19	1	3	5	0	1	74
9	TF Ria_Utari	25	44	5	56	9	14	0	5	20	0	37	10	5	7	3	47	4	4	23	266	0	54	2	107	15	3	0	9	5	2	7	0	0	19
10	TF remtaridy	32	50	7	98	25	8	2	3	21	2	48	8	2	46	9	30	4	7	54	314	0	84	0	131	21	0	2	6	3	2	14	0	0	42
11	TF shejlaputnev	32	52	14	103	22	22	4	11	28	2	55	10	3	76	19	50	2	10	37	232	1	167	0	160	29	7	0	18	10	3	8	1	0	87
12	TF vidiyhars	19	24	5	118	40	7	0	0	9	1	32	1	2	60	1	13	1	7	17	200	1	110	2	106	9	1	1	4	3	3	6	0	0	87
13	TF Verlynaemak	23	14	2	77	19	6	0	1	9	0	40	1	3	48	12	37	0	3	166	672	0	76	0	188	27	0	0	13	4	1	34	0	0	58
14	TF henfhe	11	31	5	61	10	13	1	1	10	0	32	2	3	23	2	26	1	7	23	90	2	71	0	78	12	2	1	9	15	1	9	0	0	38
15	TF nasa	16	8	1	55	13	3	0	1	0	0	8	1	0	23	6	1	0	0	6	103	0	54	0	100	11	1	0	0	0	0	0	0	0	0

Klasifikasi Logistic Regression dan TF-IDF

DISCPrediction User: Kartika Prameswari

Dashboard Control panel Home - Dashboard

Data
Crawling
[More info](#)

Term Weighting
TF-IDF
[More info](#)

Classification
Logistic Regression
[More info](#)

Pengolahan Banyak Percobaan

Pilih Metode: Keterangan Set Data:

Jumlah Data D : 19 Jumlah Data I : 119 Jumlah Data S : 182 Jumlah Data C : 134 Set Data Latih & Data Uji mengikuti

Masukan Jumlah Term Jika Term Base:

Keterangan Set Data:

Jumlah Percobaan:

Data Hasil Percobaan

id	ReD	Rel	ReS	ReC	PreD	Prel	PreS	PreC	Acc	Aksi
1	1	User Base	407	47	on	on	on	on	on	on