

Klasifikasi Kepribadian Myres-Briggs Type Indicator Menggunakan Metode TF-IDF dan Naive Bayes Classifier

¹Annisa C

Program Studi Teknik Informatika, Universitas Ngudi Waluyo
Jl. Diponegoro no 186 Gedanganak - Ungaran Timur
E-mail : annisy2302@gmail.com

²Iwan Setiawan Wibisono

Program Studi Teknik Informatika, Universitas Ngudi Waluyo
Jl. Diponegoro no 186 Gedanganak - Ungaran Timur
E-Mail: iwansetiawan@unw.ac.id

ABSTRACT

Tipe kepribadian seseorang biasanya dapat diketahui dengan menggunakan psikotes. Salah satu jenis psikotes yang sering digunakan adalah Myres-Briggs Type Indicator (MBTI). Sebuah 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. Selain itu, hasil psikotes juga dapat diketahui secara langsung. Aplikasi ini dibuat dengan menggunakan teknik text mining, yaitu dengan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) dan metode klasifikasi Naive Bayes Classifier. Metode TF-IDF digunakan untuk memberikan pembobotan kata dari data training dan metode Naive Bayes Classifier digunakan untuk mengklasifikasikan data testing berdasarkan data yang telah dilatih sebelumnya. Data akan dibersihkan terlebih dahulu sebelum dilatih maupun diuji melalui proses preprocessing sehingga informasi yang dihasilkan dapat menjadi lebih akurat. Hasil dari penelitian ini adalah dari 100 data yang diuji, 71 data termasuk kategori klasifikasi benar sesuai dengan labelnya, sedangkan 29 data termasuk kategori klasifikasi salah karena hasil prediksi tidak sesuai dengan labelnya. Tingkat akurasi adalah 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%.

Keyword : *Kepribadian, Psikotes, Myres-Briggs Type Indicator, Term Frequency-Inverse Document Frequency, Naive Bayes Classifier.*

PENDAHULUAN

Tes psikologi atau yang biasa disebut dengan psikotes merupakan salah satu cara yang dapat digunakan untuk mengetahui jenis kepribadian seseorang. Secara umum terdapat dua jenis psikotes, yaitu psikotes berbasis grafis dan psikotes berbasis kuesioner. Psikotes grafis menilai kepribadian seseorang berdasarkan gambar yang dibuatnya, sedangkan psikotes kuesioner menilai kepribadian seseorang berdasarkan jawaban yang dipilih terhadap sejumlah pertanyaan yang diberikan. Salah satu psikotes kuesioner yang sering digunakan adalah Myres-Briggs Type Indicator (MBTI). Metode ini digunakan secara spesifik untuk mempelajari dan memahami cara individu dalam menilai suatu keadaan dan mengambil keputusan atas keadaan tersebut. Katherine Cook dan putrinya yang bernama Isabel Briggs Myers adalah orang pertama yang mencetuskan metode ini, dimana metode ini merupakan pengembangan

dari teori kepribadian yang sebelumnya telah dicetuskan oleh Carl Gustav Jung [1]. MBTI seringkali digunakan sebagai alat untuk menguji kepribadian pegawai oleh perusahaan untuk menentukan perekrutan seorang karyawan. Metode ini mudah dipahami, diolah, dan bersifat aplikatif sehingga banyak digunakan dalam berbagai penelitian untuk mengidentifikasi karakter [2]. Penelitian ini menggunakan model Keirsey untuk mengelompokkan 16 jenis kepribadian yang berasal dari 4 kelompok utama yaitu dimensi pemusatan perhatian (extrovert/E vs introvert/I), dimensi memahami informasi dari luar (sensing/S vs intuition/N), dimensi menarik kesimpulan dan keputusan (thinking/T vs feeling/F), dan dimensi pola hidup (judging/J vs perceiving/P) [3]. Model tersebut dapat dilihat pada tabel 1.

Tabel 1. Model Keirsey Dari Mbti

No	Artisan	Guardian	Rational	Idealist
1	Composer	Inspector	Champion	Architect

No	Artisan (ISFP)	Guardian (ISTJ)	Rational (ENFP)	Idealist (INTP)
2	Crafter (ISTP)	Protector (ISFJ)	Counselor (INFJ)	Fieldmarshal (ENTJ)
3	Performer (ESFP)	Provider (ESFJ)	Healer (INFP)	Inventor (ENTP)
4	Promoter (ESTP)	Supervisor (ESTJ)	Teacher (ENFJ)	Mastermind (INTJ)

Pada penelitian ini, data training yang digunakan berupa sekumpulan dataset kepribadian MBTI yang telah memiliki label kelas untuk setiap data dan diambil dari situs [kaggle.com/datasnaek/mbti-type](https://www.kaggle.com/datasnaek/mbti-type) dan berjumlah 8675 baris data [4], sedangkan data testing menggunakan data dari 100 orang tokoh terkenal yang sudah diketahui label/kelas sebenarnya dengan data yang diambil dari situs [personality-databases.com](https://www.personality-databases.com). Scraping data dalam bahasa inggris diperoleh dari akun twitter para tokoh yang telah diverifikasi dan minimum memiliki pengikut lebih dari 100.000 orang.

Algoritma yang digunakan untuk melakukan klasifikasi adalah algoritma naive bayes karena algoritma naive bayes merupakan algoritma yang mudah diimplementasikan dan dilatih meskipun dengan sedikit data, sementara kekurangannya adalah terdapat sebuah kondisi yang disebut zero conditional probability problem. Pada penelitian tersebut [5], algoritma naive bayes juga dapat digunakan untuk klasifikasi teks, filter spam, analisis sentimen dan sistem rekomendasi. Apabila tingkat akurasi algoritma naive bayes dibandingkan dengan algoritma lain seperti K-Nearest Neighbor dan Support Vector Machine untuk kasus klasifikasi kepribadian Big Five Personality pada Twitter, maka naive bayes menghasilkan tingkat akurasi yang paling tinggi yaitu 60% [6]. Namun untuk klasifikasi bahasa indonesia yang berasal dari terjemahan dataset bahasa inggris, tingkat akurasi menurun dikarenakan perbedaan makna yang sebenarnya antara bahasa inggris dengan bahasa indonesia, sehingga tidak disarankan untuk melakukan penerjemahan terhadap data.

Klasifikasi kepribadian berdasarkan MBTI juga pernah diimplementasikan dengan algoritma Support Vector Machine dengan data dari akun twitter, dan menghasilkan akurasi tertinggi yaitu 88,89% dalam skenario perbandingan data training:data testing sebanyak 80%:20%. Hal ini cukup menarik karena untuk skenario 90%:10% tingkat akurasi mengalami penurunan [7]. Penelitian serupa juga dilakukan oleh Bharadwaj, Sridhar, Choudhary, dan Srinath [8] yang menyatakan bahwa algoritma SVM mengungguli naive

bayes dan neural network. Namun pada penelitian klasifikasi karakter seseorang menggunakan algoritma naive bayes dengan sumber data Twitter yang lebih sedikit, yaitu 10 data training dan 10 data testing seperti yang dilakukan oleh Sarwani dan Mahmudy (Sarwani dan Mahmudy, 2015) [9] dapat menghasilkan tingkat akurasi sebesar 100%. Tidak hanya Twitter, forum-forum online seperti StackOverflow juga dapat dijadikan sebagai sumber data untuk menganalisis kepribadian seseorang seperti yang dilakukan oleh Katiyar, Kumar, dan Walia [10]. Kepribadian ternyata juga dapat memengaruhi pencapaian akademik seseorang seperti yang dibuktikan oleh Mulyati dan Setiani [11] dengan menggunakan algoritma naive bayes, dimana data merupakan hasil kuisioner dari para responden yang telah mengikuti tes MBTI melalui website mbtiforedukasi.com. Data tersebut kemudian dibagi menjadi data training dan data testing, dimana 63 data yang berasal dari mahasiswa jurusan Psikologi angkatan 2014 digunakan sebagai data training dan 28 data yang berasal dari mahasiswa jurusan TI angkatan 2014 digunakan sebagai data testing. Tingkat akurasi yang diperoleh dari penelitian tersebut adalah sebesar 80,95%.

Oleh karena itu algoritma naive bayes dianggap mumpuni untuk digunakan dalam klasifikasi kepribadian MBTI melalui cuitan twitter. Implementasi akan menggunakan python, sehingga diharapkan dapat menjadi alternatif bagi para psikolog selain psikotes yang dilakukan secara konvensional.

METODE PENELITIAN

A. Data Pre-processing

Prapemrosesan data yang dilakukan mencakup (1) mengubah semua huruf kapital menjadi huruf kecil; (2) menghilangkan setiap kata yang memiliki unsur tautan ([http](#)), gambar ([pic](#)), tagar (<#>), dan username/tag ([@](#)); (3) menghilangkan unsur angka, dan semua jenis tanda baca dan simbol-simbol yang tidak memiliki makna, seperti seperti ["™", "©", "SM"]; (4) tokenisasi; (5) menghilangkan semua kata penghubung dan stopword, dan (6) mengembalikan kata-kata hasil tokenisasi menjadi bentuk dasar dengan cara menghilangkan awalan, akhiran, dan mengembalikan kata tersebut ke bentuk verb-1.

B. TF-IDF

Pembobotan pada setiap kata yang ada pada setiap kelas menggunakan metode Text Frequency-Inverse Document Frequency (TF-IDF). Contoh token kata hasil prapemrosesan dapat dilihat pada tabel 2.

Tabel 2. Contoh Token Data

	Artisan	Guardian	Rational	Idealist
Token Kalimat1	[paint, number, guess, entertain, bad]	[speak, white, man, black, amaze]	[nation, soldier, independence, soldier, soldier]	[amaze, car, guess, bad, amaze]
Token Kalimat 2	[fire, soldier, black, white, fire]	[cat, entertain, bad, cat, bad]	[country, big, small, thunder, big]	[chess, art, play, decision, nation]

Langkah-langkah penyelesaiannya adalah:

- 1) Menghitung perbandingan antara jumlah kata tertentu dengan jumlah kata keseluruhan dalam 1 kalimat (Text Frequency).
- 2) Menghitung perbandingan antara jumlah kata tertentu yang muncul pada kalimat tertentu dengan jumlah semua kalimat yang ada dalam 1 dokumen (Document Frequency).
- 3) Menghitung invers dari Document Frequency sehingga menghasilkan Inverse Document Frequency.
- 4) Mengalikan Text Frequency dengan Inverse Document Frequency sehingga mendapatkan Term Frequency-Inverse Document Frequency.
- 5) Membuat kamus data dari hasil perhitungan TF-IDF, dimana token kalimat 1 dan token kalimat 2 digabungkan. Kata yang sama pada token kalimat 1 dan token kalimat 2 akan dijumlahkan nilai TF-IDF nya. Contoh hasil akhir perhitungan TF-IDF untuk kelas artisan dapat dilihat pada tabel 3.

Tabel 3. Contoh Perhitungan Tf-IDF Kelas Artisan

	Kata	Nilai TF-IDF
Token Kalimat 1	paint	$0,2 \times 0,301029995 = 0,060205999$
	number	$0,2 \times 0,301029995 = 0,060205999$
	guess	$0,2 \times 0,301029995 = 0,060205999$
	entertain	$0,4 \times 0,301029995 = 0,120411998$
Token Kalimat 2	fire	$0,4 \times 0,301029995 = 0,120411998$
	soldier	$0,2 \times 0,301029995 = 0,060205999$
	black	$0,2 \times 0,301029995 = 0,060205999$
	white	$0,2 \times 0,301029995 = 0,060205999$

C. Naive Bayes Classifier

Metode Naive Bayes Classifier digunakan untuk melakukan klasifikasi pada data testing dengan cara mengelompokkan data testing tersebut ke salah satu kelas Artisan, Guardian, Rasional, atau Idealis. Langkah-langkah pengerjaannya sebagai berikut:

- 1) Membagi data testing yang terkumpul sesuai dengan keyword user id Twitter yang dimasukkan menjadi kalimat-kalimat terpisah.
- 2) Memecah kalimat menjadi kata dan kemudian dilakukan pre-processing.
- 3) Memetakan probabilitas setiap kata dengan cara mengambil probabilitas setiap kata yang sama pada data testing dengan yang terdapat di dalam setiap kelas data training. Sebagai contoh kata "work" pada kelas Artisan memiliki probabilitas sebesar W%, pada kelas Guardian memiliki probabilitas sebesar X%, pada kelas Rasional memiliki probabilitas sebesar Y%, dan pada kelas Idealis memiliki probabilitas sebesar Z%, maka pada data testing yang mengandung kata "work" akan memiliki 4 kemungkinan.
- 4) Menghitung probabilitas kata dalam data testing yang tidak terdapat pada data training menggunakan Laplace Smoothing.
- 5) Mengalikan probabilitas dari setiap kata per kalimat sesuai dengan probabilitas yang telah dipetakan.
- 6) Membandingkan hasil perkalian antara 4 kelas dan hasil dengan probabilitas terbesar akan menjadi kelas prediksi dari data testing tersebut.

Contoh proses klasifikasi data testing:

- 1) Diketahui token data testing yang akan diklasifikasi:

Token data testing = ["soldier", "country", "nation"]

- 2) Mencari probabilitas dari data testing terhadap masing-masing kelas seperti yang ditunjukkan pada tabel 4. Contoh perhitungan kelas artisan adalah dengan notasi sebagai berikut:

$$P(\text{data testing} | \text{Artisan}) = P(\text{soldier} | \text{Artisan}) \times P(\text{country} | \text{Artisan}) \times P(\text{nation} | \text{Artisan}) \times P(\text{Artisan})$$

dimana:

- a. $P(\text{soldier} | \text{Artisan})$ merupakan nilai TF-IDF kata "soldier" pada kamus data Artisan yaitu 0,06205999.
- b. $P(\text{country} | \text{Artisan})$, karena tidak terdapat kata country pada kamus data Artisan maka: $P(\text{country} | \text{Artisan}) = (0+1)/(8+25) = 0,03030$
- c. $P(\text{nation} | \text{Artisan})$, karena tidak terdapat kata nation pada kamus data Artisan maka bobotnya juga adalah 0,03030, sehingga: $P(\text{data testing} | \text{Artisan}) = 0,06205999 \times 0,030303030 \times 0,030303030 \times 0,25$
 $P(\text{data testing} | \text{Artisan}) = 1,382 \times 10^{-5}$

- d. Pada tabel 4, nilai naïve bayes classifier kelas rational adalah $1,637 \times 10^{-4}$, maka token data testing = ["soldier", "country", "nation"] diklasifikasikan sebagai kelas rational.

Tabel 4. Hasil Klasifikasi Naïve Bayes Classifier

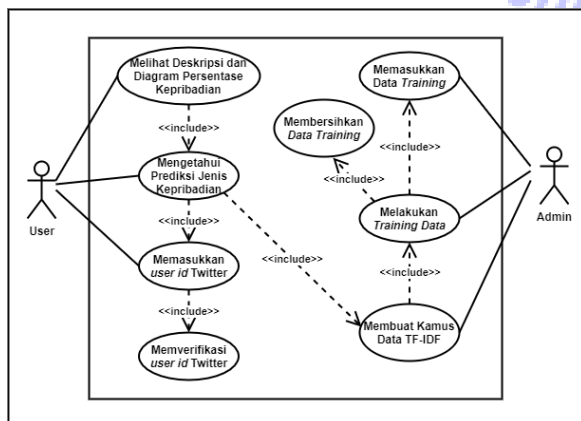
Data testing	Artisan	Guardian	Rational	Idealist
soldier	0,060205999	0,030303030	0,180617997	0,030303030
country	0,030303030	0,030303030	0,060205999	0,030303030
nation	0,030303030	0,030303030	0,060205999	0,060205999
Likehood	$5,5286 \times 10^{-5}$	$2,7826 \times 10^{-5}$	$9,5470 \times 10^{-4}$	$5,5286 \times 10^{-5}$
Nilai Prior	0,25	0,25	0,25	0,25
Naive Bayes	$1,382 \times 10^{-5}$	$6,957 \times 10^{-6}$	$1,637 \times 10^{-4}$	$1,382 \times 10^{-5}$

HASIL DAN PEMBAHASAN

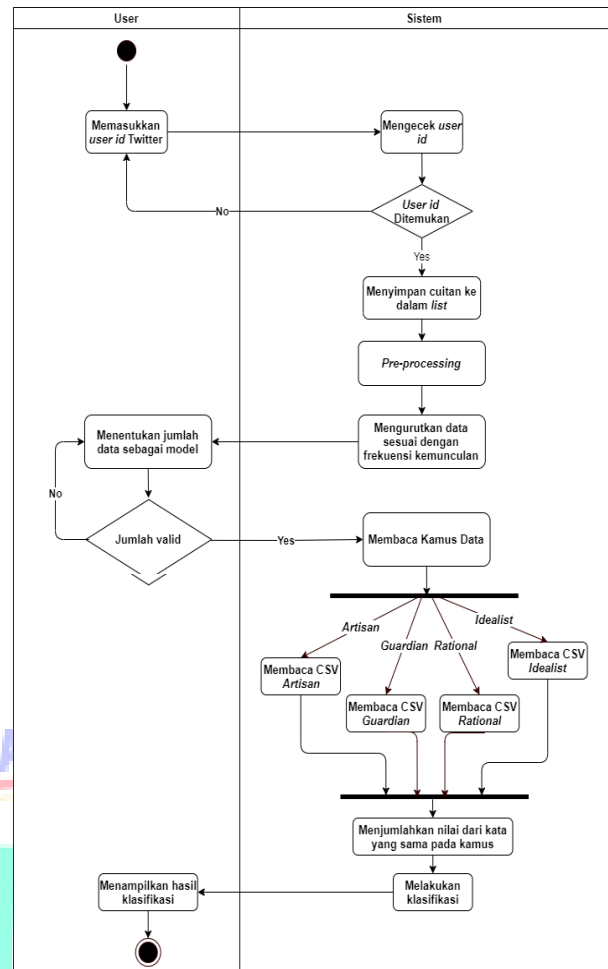
Perancangan proses mencakup *flowchart* untuk proses prapemrosesan, perhitungan TF-IDF dan klasifikasi naïve bayes classifier, sedangkan perancangan sistem menggunakan usecase diagram, activity diagram dan sequence diagram. Sistem ini diimplementasikan menggunakan bahasa pemrograman *python*. Kode program dibagi menjadi 4 bagian, yaitu proses pengelompokan data, proses training data, proses yang dilakukan pada bagian *back-end* yaitu proses uji data/klasifikasi, dan *front-end* untuk tampilan *user interface*.

A. Perancangan

Use case diagram digunakan untuk menggambarkan interaksi yang terjadi antara user dan sistem, seperti yang ditunjukkan pada gambar 1, sedangkan *activity diagram* untuk proses klasifikasi data testing yang ditunjukkan pada gambar 2 menunjukkan bahwa sistem akan melakukan verifikasi user id twitter terlebih dahulu.



Gambar 1. Use case diagram
[Sumber : Penulis, 2024]

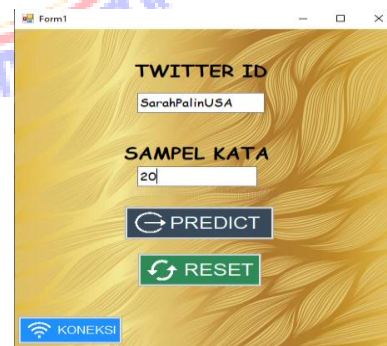


Gambar 2. Activity diagram klasifikasi data testing

[Sumber : Penulis, 2024]

B. Implementasi

Implementasi user interface dari proses ini dapat dilihat pada gambar 3.



Gambar 3. Halaman Utama
[Sumber : Penulis, 2024]

Proses selanjutnya adalah melakukan scraping terhadap data cuitan pada akun twitter tersebut. Hasil scraping akan disimpan dalam suatu variable dengan tipe data list, seperti yang dapat dilihat pada gambar 4.


```

No                               Cuitan
0  https://www.governorpalin.org/2020/03/30/vira...
1  https://www.nbcnews.com/news/amp/ncna1164696 ...
2  Our hearts go out to the friends, family and N...
3  Death Panels. See what I meant? So what if you...
4  Figured I'd do it because, well, this is Ameri...
..
95 Sweet Freedom Daily Devotional: Hope In Things...
96 MSNBC's Chris Hayes Mocks People Who Were Conc...
97 CNN's April Ryan Fumes, Says To Gov. Huckabee:...
98 Biden's First #MeToo Accuser Comes Back, RIPS ...
99 Tucker Describes the 'Single Most Humiliating ...

```

Gambar 4. Kalimat hasil scraping

Selanjutnya dilakukan prapemrosesan untuk data testing. Contoh kata-kata dengan frekuensi tertinggi untuk akun twitter “SarahPalinUSA” dapat dilihat pada gambar 5. Kata-kata tersebut kemudian digunakan untuk perhitungan dalam metode naive bayes classifier, sementara kata-kata lainnya tidak akan digunakan. Dari hasil perhitungan yang ditunjukkan pada gambar 6, maka akan diperoleh probabilitas posterior naive bayes untuk kelas artisan adalah $5,5 \times 10^{-67}$, paling besar dibandingkan kelas lainnya, sehingga akun twitter “SarahPalinUSA” diklasifikasikan sebagai kelas artisan. Tampilan hasil klasifikasi kepribadian tersebut dapat dilihat pada gambar 7. Gambar tersebut juga menunjukkan persentase kelas dalam bentuk radar chart, yaitu untuk kelas artisan sebesar 67%, dan untuk kelas-kelas lainnya, yaitu kelas idealist sekitar 20%, kelas rasional sekitar 12%, dan kelas guardian sebesar 1%.

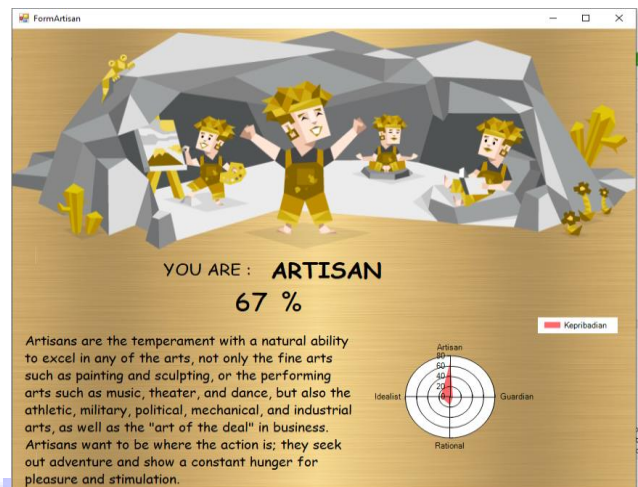
Word	Frequency
0	love 5
1	god 5
2	family 4
3	media 4
4	check 4
5	video 4
6	larry 3
7	work 3
8	great 3
9	sarah 3
10	legends 3
11	sturgis 3
12	bill 3
13	campaign 2
14	death 2
15	rule 2
16	reality 2
17	show 2
18	big 2
19	stage 2

Gambar 5. Kata dengan frekuensi tertinggi

```

Artisan : [0.029053977918277213, 0.026026125558772508, 0.04580552155020251, 6.131056281919505e-09, 0.02659424403907194, 0.03
3162541761603106, 6.131056281919505e-09, 0.031470436572620185, 0.03720072085798367, 6.131056281919505e-09, 6.131056281919505e
-09, 6.131056281919505e-09, 0.008166426631661065, 0.006592202914750924, 0.01991950674932467, 0.017306878099794025, 0.01802804
4524095844, 0.042541772401542076, 0.03154520039640824, 0.003955369748858554]
Jumlah : 5.502632963432782e-67

```

Gambar 6. Hasil Naïve Bayes Kelas Artisan
[Sumber : Penulis, 2024]Gambar 7. Hasil klasifikasi
[Sumber : Penulis, 2024]

C. Pengujian

Pengujian menggunakan confusion matrix seperti tabel 5. Proses pengujian ini mencakup tingkat akurasi, precision dan recall. Masing-masing confusion matrix data training dan data testing ditunjukkan oleh tabel 6 dan 7.

Tabel 5. Pengujian Dengan Confusion Matrix

		Kelas Prediksi			
Kelas Sebenarnya		Artisan	Guardian	Idealist	Rational
	Artisan	TP _A	E _{AB}	E _{AC}	E _{AD}
	Guardian	E _{BA}	TP _B	E _{BC}	E _{BD}
	Idealist	E _{CA}	E _{CB}	TP _C	E _{CD}
	Rational	E _{DA}	E _{DB}	E _{DC}	TP _D

Tabel 6. onfusion Matrix Data Training

Prediksi aktual	Artisan	Guardian	Rational	Idealist	Jumlah
Artisan	55	0	0	0	55
Guardian	2	48	0	0	50
Rational	0	1	47	0	48
Idealist	1	2	0	47	50
Jumlah	58	51	47	47	203

Pada tabel 5, dapat dilihat bahwa tingkat akurasi data training adalah sebesar 97%. Selanjutnya, untuk data testing akan dilakukan pengujian tingkat akurasi, nilai precision dan recall untuk setiap kelas.

Tabel 7. Confusion Matrix Data Testing

Prediksi aktual \	Artisan	Guardian	Rational	Idealist	Jumlah
Artisan	17	2	4	2	25
Guardian	4	17	4	0	25
Rational	2	3	18	2	25
Idealist	4	0	2	19	25
Jumlah	28	21	28	23	100

Tingkat akurasi pada data testing adalah sebesar 71%, sedangkan nilai precision dan recall untuk masing-masing kelas adalah sebagai berikut:

1. Kelas Artisan, nilai precision sebesar 63% dan recall sebesar 68%.
2. Kelas Guardian, nilai precision sebesar 77% dan recall sebesar 68%.
3. Kelas Rational, nilai precision sebesar 64% dan recall sebesar 72%.
4. Kelas Idealist, nilai precision sebesar 83% dan recall sebesar 76%.

Dengan demikian maka dapat dilihat bahwa naive bayes classifier cukup menggambarkan kemiripan/ kedekatan antara kelas prediksi dan kelas sebenarnya, ketepatan antara kelas prediksi dengan label yang sudah ditentukan dibandingkan dengan total semua kelas yang berhasil diprediksi oleh sistem, dan cukup berhasil mengidentifikasi semua data yang merupakan anggota dari kelas tertentu.

KESIMPULAN

Metode TF-IDF yang diimplementasikan untuk melakukan pembobotan terhadap setiap kata dengan naive bayes classifier dapat diterapkan untuk kasus klasifikasi kepribadian berdasarkan cuitan di twitter, meskipun TFIDF menghasilkan kata-kata dengan bobot terbesar yang berbeda antar-kelas artisan, guardian, rational, maupun idealist. Penelitian selanjutnya dapat menggunakan data training dari para ahli psikologi secara langsung. Selain itu, teknik prapemrosesan dapat diperkaya dengan penambahan stopword, kombinasi stemming dan lemmatization sehingga dapat dihasilkan kata-kata yang lebih mendekati bentuk kata dasar. Metode n-gram juga dapat diterapkan sehingga hasil klasifikasi dapat mencakup makna kata dalam sebuah konteks kalimat.

UCAPAN TERIMA KASIH

Ucapan terimakasih kepada Seluruh Dosen dan staf kependidikan Universitas Ngudi Waluyo atas dukungan moril dalam terlaksananya penelitian yang penulis lakukan, serta teman sejawat dalam suport dan doa sehingga penelitian dapat terselesaikan dengan baik.

DAFTAR PUSTAKA

1. Mely Amaliyah, Fiftin Noviyanto. 2013. Aplikasi Tes Kepribadian untuk Penempatan Karyawan Menggunakan Metode MBTI (Myres-Briggs Type Indicator) Berbasis Web (Studi Kasus : PT .Winata Putra Mandiri). Jurnal Sarjana Teknik Informatika, 1(2). hal.607-616.
2. Muhammad Fatroni, Erlin. 2015. Aplikasi Menentukan Karakter Peserta Didik Menggunakan Teori Myres-Briggs Type Indicator. SATIN-Sains dan Teknologi Informasi, 1(2). hal.41-47.
3. David Keirse. The Four Temperament. <https://keirse.com/temperament-overview>. Diakses tanggal 10 Februari 2020
4. Mitchell J. 2018. (MBTI) Myress-Briggs Personality Type Dataset. <https://www.kaggle.com/datasnaek/mbti-type>. Diakses tanggal 10 Februari 2020.
5. Pouria Kaviani, Sunita Dhorte. 2017. A Short Survey On Naive Bayes Algorithm. International Journal of Advance Engineering and research Development, 4(11). hal.607-611.
6. Bayu Yudha Pratama, Riyanarto Sarno. 2015. Personality Classification Based On Twitter Text Using Naive Bayes, KNN, and SVM. 2015 International Conference on Data and Software Engineering, 25-26 November 2015, Yogyakarta, Indonesia. hal.170-174.
7. Muhammad Fikry, Yusra. 2018. Ekstrover atau Introver : Klasifikasi Kepribadian Pengguna Twitter Dengan Menggunakan Metode Support Vector Machine. Jurnal Sains, Teknologi, dan Industri, 16(1). hal.72-76.
8. Srilakshmi Bharadwaj, dkk. 2018. Persono Traits Identification Based On Myres-Briggs Type Indicator (MBTI) – A Text Classification Approach. 2018 International Conference on Advances in Computing, Communications and Informatics, 19-22 September 2018, Banglore, India. hal.1076-1082.
9. Mohammad Zoqi Sarwani, Wayan Firdaus Mahmudy. Analisis Twitter Untuk Mengetahui Karakter Seseorang Menggunakan Algoritma Naive Bayess Classifier. Seminar Nasional Sistem Informasi Indonesia, 2-3 November 2015, Surabaya, Indonesia. hal. 291-296.
10. Sandhya Katiyar, Sanjay Kumar, Himdweep Walia. 2020. Personality Prediction From Stack Overflow By Using Naive Bayess Theorem In Data Mining. International Journal of Innovative Technology and Exploring Engineering, 9(3). hal.1555-1559.

11.Sri Mulyati, Novi Setiani. 2018. Identifying Students Academic Achievement And Personality Type with Naive Bayess

Classification. Jurnal Sekolah Tinggi Manajemen Informatika dan Komputer, 22(2). hal.64-68.

