

SISTEM PREDIKSI KEPRIBADIAN MANUSIA BERDASARKAN STATUS MEDIA SOSIAL MENGGUNAKAN SUPPORT VECTOR MACHINE

Elsen Ronando^{1,*}, Muhammad Yasa², Enny Indasyah³

^{1,2}Teknik Informatika, Fakultas Teknik, Universitas 17 Agustus 1945 Surabaya

³Teknik Elektro Otomasi, Fakultas Vokasi, Institut Teknologi Sepuluh Nopember
Surabaya

Email: *elsen.ronando@untag-sby.ac.id

ABSTRACT

Currently, social media is a forum for exchanging information widely used by the public, such as Facebook and Twitter. Social media users exchange information to find out the condition of one another. Some companies use social media to explore the personality potential of prospective employees to be recruited. However, to dig up this information takes a very long time because the company has to open prospective employees' social media one by one. To dig up information automatically, a personality detection system is needed from social media users. This study develops a person's personality prediction system based on social media status using the support vector machine. The data sets evaluated in this study were 300 Facebook social media status data and 2067 Twitter social media status data. Based on the evaluation results, we obtained a high level of accuracy in detecting a person's personality based on social media status, namely 100% for Facebook user status and 99.3% for Twitter user status.

Keywords: Personality, Social Media, Support Vector Machine, Facebook, Twitter.

ABSTRAK

Saat ini, media sosial merupakan salah satu wadah pertukaran informasi yang banyak digunakan oleh masyarakat, seperti Facebook maupun Twitter. Pengguna media sosial saling bertukar informasi untuk mengetahui kondisi satu dengan lainnya. Beberapa perusahaan memanfaatkan media sosial untuk menggali potensi kepribadian dari calon pegawai yang akan direkrut. Namun, untuk menggali informasi tersebut memerlukan waktu yang sangat lama karena perusahaan harus membuka media sosial dari calon pegawai satu per satu. Agar dapat menggali informasi secara otomatis, maka diperlukan sistem deteksi kepribadian dari pengguna media sosial. Penelitian ini mengembangkan sistem prediksi kepribadian seseorang berdasarkan status media sosial menggunakan metode Support Vector Machine. Set data yang dievaluasi dalam penelitian ini yaitu 300 data status media sosial Facebook dan 2067 data status media sosial Twitter. Berdasarkan hasil evaluasi yang dilakukan diperoleh tingkat akurasi yang tinggi dalam mendeteksi kepribadian seseorang berdasarkan status media sosial, yaitu 100% untuk status pengguna Facebook dan 99,3% untuk status pengguna Twitter.

Kata Kunci: Kepribadian, Media Sosial, Support Vector Machine, Facebook, Twitter.

1. Pendahuluan

Media sosial merupakan sebuah wadah yang banyak digunakan masyarakat saat ini untuk saling bertukar informasi [1], [2]. Facebook dan Twitter adalah perangkat media

sosial yang populer digunakan masyarakat umum, disebabkan kemudahan dalam penggunaannya.

Pemanfaatan media sosial ini juga banyak digunakan oleh perusahaan dalam proses penjangkaran

calon pegawai baru. Beberapa perusahaan dalam seleksi calon pegawai baru selalu menyertakan *form* nama media sosial untuk diisi oleh calon pegawai, baik Facebook maupun Twitter. Tujuan pemanfaatan media sosial dalam proses seleksi ini adalah bertujuan untuk menganalisa kepribadian dari calon pegawai berdasarkan postingan status dari calon pegawai. Kepribadian yang sesuai dengan perusahaan menjadi bahan pertimbangan calon pegawai tersebut diterima dalam perusahaan.

Namun, pada kenyataannya, perusahaan dengan jumlah pelamar yang banyak memerlukan waktu yang lama untuk menganalisa kepribadian calon pegawai berdasarkan status Facebook maupun Twitter. Oleh sebab itu, dibutuhkan sistem yang dapat mengatasi permasalahan tersebut.

Machine learning atau yang biasa disebut juga dengan pembelajaran mesin adalah metode yang banyak digunakan dalam menyelesaikan permasalahan analisa prediktif, seperti analisa kepribadian manusia [3]–[7]. Salah satu penelitian yang menggunakan metode pembelajaran mesin adalah klasifikasi kepribadian berdasarkan status Facebook dengan menggunakan metode *backpropagation* dengan tingkat akurasi 84% [8]. Namun, penelitian tersebut masih memiliki kelemahan, yaitu belum dapat menentukan hasil yang optimal pada konfigurasi *neuron* tertentu. Penelitian lain untuk klasifikasi kepribadian berdasarkan status Facebook menggunakan metode *naïve Bayes* dengan akurasi 59,9% [9]. Berdasarkan penelitian yang telah dilakukan menunjukkan bahwa masih diperlukan sebuah sistem yang dapat meningkatkan akurasi yang lebih baik. Metode Support Vector Machine untuk

klasifikasi kepribadian berdasarkan media sosial Twitter juga telah dilakukan oleh [10], dengan tingkat akurasi 86,88%. Selain itu, metode Support Vector Machine ini juga memiliki waktu komputasi yang optimal [11].

Berdasarkan hal tersebut, penelitian ini berfokus pada sistem prediksi kepribadian manusia berdasarkan status media sosial, baik Facebook maupun Twitter menggunakan metode Support Vector Machine. Metode Support Vector Machine ini diterapkan karena memiliki performansi yang lebih dalam tingkat akurasi daripada beberapa metode yang lain, seperti *backpropagation* maupun *naïve Bayes*.

2. Tinjauan Pustaka

Bagian ini menjelaskan landasan teori yang digunakan dalam penelitian ini.

2.1. *Linguistic Inquiry and Word Count (LIWC)*

LIWC adalah sebuah program analisis teks psikologi dalam digital yang dikembangkan oleh James W. Pennebaker dan Martha E. Francis [12]. LIWC sendiri bertujuan untuk menganalisa psikologi dalam teks atau kalimat digital.

Saat ini, dengan menggunakan metode analisis teks LIWC dapat diperoleh informasi tentang psikologi seseorang. Cara kerja LIWC pada dasarnya sangat sederhana, karena hanya membaca teks yang diberikan dan menghitung prosentase kata-kata yang menggambarkan emosi yang berbeda, gaya berpikir dan kepedulian terhadap sosial.

Terdapat beberapa kategori yang dianalisa oleh LIWC untuk mengidentifikasi sebuah teks, antara lain *summary language variables*,

linguistic dimensions, other grammar, dan psychological process.

2.2. Kepribadian

Kepribadian adalah suatu sikap dari seseorang sebagai latar belakang perilaku yang dilakukannya.

Kepribadian secara umum terdapat 8 (delapan) pribadi antara lain senang, sedih, terkejut, takut, perasaan tidak senang, marah, khawatir, dan percaya. Berdasarkan hal tersebut, kepribadian dapat digolongkan dalam tiga, yaitu *negative emotion, positive emotion, dan social process* [13].

Kepribadian *negative emotion* adalah *anxiety* (gelisah), *anger* (marah), dan *sadness* (sedih). Kepribadian *positive emotion* adalah *love* (cinta), *nice* (perasaan bagus), dan *sweet* (perasaan manis). Kepribadian *social process* adalah *family* (rasa kekeluargaan), *friend* (teman), *female references* (rasa keibuaan), dan *male references* (rasa kebabakan).

2.3. Support Vector Machine

Support Vector Machine (SVM) merupakan metode pembelajaran yang bertujuan untuk mengklasifikasikan data dalam dua maupun lebih kelompok, baik menggunakan fungsi-fungsi *linear* maupun *non-linear*. Klasifikasi SVM memiliki tujuan mencari *hyperplane* (garis potong) terbaik dengan margin yang optimal. Margin merupakan ukuran jarak antara *hyperplane* dengan titik terdekat pada setiap kelas, yang disebut dengan *support vector* [14].

Gambar 1 menunjukkan terdapat *hyperplane* (garis potong) yang memisahkan dua kelas secara linier. Pada bagian batas awal membatasi kelas pertama ($y_i = 1$), dan bagian batas kedua membatasi kelas kedua ($y_i = -1$). Kemudian, data tersebut dikalkulasi berdasar perkalian vektor,

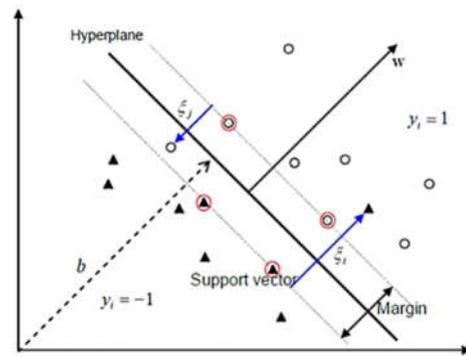
yaitu vektor bobot dengan set data seperti ditunjukkan pada Persamaan (1) dan (2):

$$w \cdot x_i + \geq b \tag{1}$$

untuk kelas 1 $y_i = 1$

$$w \cdot x_i + \leq b \tag{2}$$

untuk kelas 2 $y_i = -1$, dimana x_i set data, y_i kelas data x_i , w vektor bobot yang tegak lurus dengan garis potong, dan b fungsi pemisah terhadap titik asal.



Gambar 1. Support Vector Machine (SVM) secara linier

Garis potong optimal merupakan sebuah bidang pemisah dengan margin maksimal atau memiliki posisi tengah diantara dua kelas yang berbeda. Nilai margin antara dua bidang pembatas dapat dihitung dengan $m = \frac{2}{\|w\|}$. Sementara itu, untuk memperoleh margin maksimal dapat dihitung dengan fungsi Lagrangian pada Persamaan (3):

$$\min_{w,b,L_p} (w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^1 (a_i y_i ((w \cdot x + b) - 1)) + \sum_{i=1}^1 a_i \tag{3}$$

Dimana a_i adalah Lagrange multiplier yang berhubungan dengan x . Vektor w pada umumnya bernilai sangat besar sampai tak terhingga. Agar nilai tersebut menjadi terhingga, maka persamaan *Lagrange primal problem*

tersebut dikonversi kedalam *Lagrange dual problem* seperti Persamaan (4) :

$$L_p(a) = \sum_{i=1}^1 a_i - \frac{1}{2} \sum_{j=1}^1 a_i a_j y_i y_j x_i x_j \quad (4)$$

Karena $\min_{w,b} L_p = \max_a L_d$. Maka nilai pencarian *hyperplane* terbaik dapat diformulasikan pada Persamaan (5) dan (6):

$$\text{Max}_a L = \sum_{i=1}^1 a_i - \frac{1}{2} \sum_{i=1}^1 \sum_{j=1}^1 a_i a_j y_i y_j x_i x_j \quad (5)$$

$$s. t \sum_{i=1}^1 a_i y_i = 0 \geq 0 \quad (6)$$

Persamaan (6) memperoleh nilai a_i pada setiap model data. Kemudian nilai yang diperoleh digunakan dalam penentuan bobot (w). Data dengan nilai $a_i \geq 0$ maka disebut *support vector*, sementara sisanya bukan *support vector*. Setelah a_i diperoleh, kemudian kelas data uji x_i maka dapat diformulasikan pada Persamaan (7) :

$$f(x_d) = \sum_{i=1}^{NS} a_i y_i x_i x_d + b \quad (7)$$

dimana x_i *support vector*, NS jumlah *support vector*, x_d data klasifikasi.

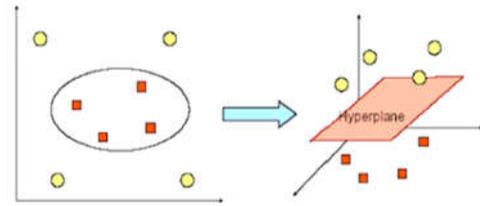
Kemudian pada ruang vektor baru tersebut, dicari garis potong terbaik yang bertujuan memisahkan dua kelas secara *linear*. Pencarian garis potong tergantung pada *dot product* dari pemetaan ruang dimensi tinggi, yaitu $\Phi(u)\Phi(v)$, seperti pada Gambar 2. Nilai hitung dari transformasi $\Phi(u)$ memiliki tingkat kompleksitas, sehingga dapat digantikan dengan fungsi *kernel* sesuai pada Persamaan (8) :

$$K(u,v) = \Phi(u) \cdot \Phi(v) \quad (8)$$

Jika Persamaan (8) disubstitusi dalam persamaan *Lagrangian* maka diperoleh Persamaan (9) :

$$L = \sum_{i=1}^1 a_i - \frac{1}{2} \sum_{j=1}^1 a_i a_j y_i y_j K(u,v) \quad (9)$$

dengan $f(x_d) = \sum_{i=1}^{NS} a_i y_i \cdot K(u,v) + b$ dimana u data pelatihan, v kelas data pelatihan, d dan γ parameter *kernel*.



Gambar 2. Support Vector Machine (SVM) Secara non- linear

3. Metode

Berikut ini merupakan tahapan penelitian yang dilakukan sebagai berikut.

3.1. Pengumpulan Data

Pada tahap ini dilakukan proses pengumpulan data berupa data status dari 30 pengguna Facebook dengan batasan usia antara umur 20 sampai 30 tahun dan status yang diambil menggunakan bahasa inggris serta setiap pengguna diambil masing-masing 10 status, sehingga mendapatkan 300 status pengguna Facebook. Metode yang digunakan untuk mengambil data status Facebook adalah metode *crawling* [15]. Kemudian diambil juga data pengguna Twitter dengan cara *scraping* [16] menggunakan 3 kalimat yaitu *Country, Friends, High School* hingga didapatkan data sebanyak 2067 status *tweets*.

3.2. Pengolahan Data

Pada tahap ini data status Facebook maupun Twitter yang telah dikumpulkan, kemudian diolah menggunakan program LIWC seperti yang ditunjukkan pada Gambar 3. Tujuannya adalah untuk menganalisa nilai kepribadian seseorang

berdasarkan makna dari status kalimat. Kemudian, dilakukan seleksi data dari nilai yang dihasilkan LIWC dengan tujuan untuk memilih atribut yang digunakan dalam mendeteksi kepribadian seseorang. Atribut atau fitur yang dipilih terdiri dari 3 (tiga) atribut, yaitu atribut posemo (emosi positif), negemo (emosi negatif), dan sosial (emosi sosial), seperti yang ditunjukkan pada Tabel 1. Selanjutnya, set data dibagi menjadi 80% untuk data latih dan 20% untuk data uji, seperti yang ditunjukkan pada Tabel 1. Pembagian data ini didasarkan pada penelitian yang dilakukan oleh [17], dimana secara empirik dengan data latih 80% dan data uji 20% memiliki hasil yang prediksi yang maksimal.

Dalam pengolahan data, data 300 status Facebook dilakukan normalisasi dengan cara menghapus data yang

sama sehingga didapatkan 204 data status, dari 204 data status dibagi menjadi 2 bagian, yang pertama 159 data status pengguna dipergunakan untuk data latih dan bagian yang kedua 45 untuk data uji. Kemudian status data dari Twitter data *tweets* dinormalisasi dan dibagi 1653 untuk data latih dan 414 untuk data uji.

Tabel 1. Nilai Status Media Sosial Menggunakan LWC

Status	Emosi Positif	Emosi Negatif	Sosial	Hasil
I lost in a tournament!	0,00	20,00	0,00	Emosi Negatif
With hahaha hill	28,57	0,00	0,00	Emosi Positif
⋮	⋮	⋮	⋮	⋮

Source (A)	LWC	Analytic	Clout	AuthenticTone	WPS	Sixtr	Dic	function	pronoun	ppron	i	we	you	shehe	they	ipron	article	p
Joseph Yap	2	93,26	50,00	1,00	25,77	2,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Call It What You Want	12	6,95	99,00	1,00	25,77	12,00	8,35	91,67	41,67	16,67	0,00	16,67	0,00	0,00	0,00	0,00	25,00	0,00
so who has finish the black scholes part?????	8	11,86	89,42	7,84	25,77	8,00	12,50	87,50	50,00	12,50	0,00	0,00	0,00	0,00	0,00	0,00	12,50	12,50
Final Sem seems like a real joy to watch..... Hmmmmm.....	10	99,00	50,00	43,37	99,00	5,00	0,00	80,00	30,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	10,00
chelsea, how many yellow cards do u still wanna get, you are	16	1,92	99,00	7,84	25,77	16,00	12,50	87,50	56,25	12,50	12,50	0,00	0,00	12,50	0,00	0,00	0,00	0,00
went down my house to play soccer, honestly, not impressed	16	97,36	10,58	95,60	99,00	16,00	12,50	93,75	43,75	6,25	6,25	6,25	0,00	0,00	0,00	0,00	0,00	6,25
Koreans copy only bleh~~~~	4	38,60	50,00	1,00	25,77	4,00	25,00	25,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
So who is faster? bike or horse??	7	1,00	50,00	89,63	25,77	3,50	0,00	85,71	57,14	14,29	0,00	0,00	0,00	0,00	0,00	0,00	14,29	0,00
Lost my Phone AND Sim card.... Anyone who CAN BE BOTHERE	35	3,86	95,65	82,17	1,91	11,67	8,57	85,71	60,00	28,57	20,00	14,29	0,00	5,71	0,00	0,00	8,57	2,86
After today, i can finally say BYE to late night working!!!! Lar	14	81,81	76,25	99,00	25,77	7,00	14,29	85,71	42,86	7,14	7,14	7,14	0,00	0,00	0,00	0,00	0,00	0,00
Weekends seems fun uh??? NOOO!!! definitely not when you r	22	69,94	50,00	96,40	99,00	7,33	18,18	95,45	54,55	9,09	4,55	0,00	0,00	4,55	0,00	0,00	4,55	9,09
Barlow Everett	2	93,26	50,00	1,00	25,77	2,00	50,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Women are the joy and pillar of life..!	8	93,26	89,42	1,00	99,00	8,00	0,00	87,50	50,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	12,50
Inbox for more info about these products. Email.. info@arene	11	93,26	81,84	2,40	25,77	3,67	18,18	63,64	27,27	9,09	0,00	0,00	0,00	0,00	0,00	0,00	9,09	0,00
New year offer. chat me on whats app.. +1(614)3838769	11	57,84	81,84	99,00	25,77	3,67	0,00	72,73	27,27	18,18	9,09	9,09	0,00	0,00	0,00	0,00	9,09	0,00
To become a man is not a day job	9	93,26	13,32	98,01	25,77	9,00	0,00	100,00	66,67	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	22,22
love is like a fire, but whether it is going to warm your heart	23	15,15	95,89	80,16	99,00	23,00	4,35	100,00	65,22	17,39	13,04	0,00	0,00	13,04	0,00	0,00	4,35	4,35
You Asked me what was wrong and i smiled and said..NOTH	19	1,00	70,08	96,09	25,77	19,00	15,79	94,74	68,42	31,58	21,05	15,79	0,00	5,26	0,00	0,00	10,53	0,00
we learn some thing 4m prosperty we learn many more 4m	12	38,60	99,00	23,51	1,00	12,00	16,67	75,00	25,00	25,00	16,67	0,00	16,67	0,00	0,00	0,00	8,33	0,00
my black berry is on d way	7	68,29	7,67	99,00	25,77	7,00	0,00	71,43	42,86	14,29	14,29	14,29	0,00	0,00	0,00	0,00	0,00	0,00
am d happiest girl in d planet	7	93,26	92,33	13,15	25,77	7,00	14,29	42,86	28,57	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
i gess olathey Yusuf akande his d most happies man in d pla	13	82,82	77,92	1,40	25,77	13,00	7,69	38,46	23,08	15,38	15,38	7,69	0,00	0,00	0,00	7,69	0,00	0,00
K&T Lim	2	93,26	50,00	1,00	25,77	2,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Happy Birthday legend i ????	3	93,26	50,00	98,01	99,00	3,00	33,33	66,67	40,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Back in track i Done workout alone ?? #teamTR #taijping	14	83,81	76,25	89,63	1,00	7,00	50,00	71,43	21,43	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Legend is back on fire ?????? #NotHumansLimited	6	61,04	50,00	98,01	25,77	6,00	16,67	66,67	50,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Stay tuned i ??????	2	93,26	50,00	99,00	25,77	2,00	0,00	50,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Yesterday is gone. Tomorrow has not yet come. We have onl	15	1,00	90,87	99,00	25,77	3,75	13,33	100,00	60,00	13,33	13,33	0,00	13,33	0,00	0,00	0,00	0,00	0,00

Gambar 3. Hasil perhitungan status media sosial menggunakan LIWC

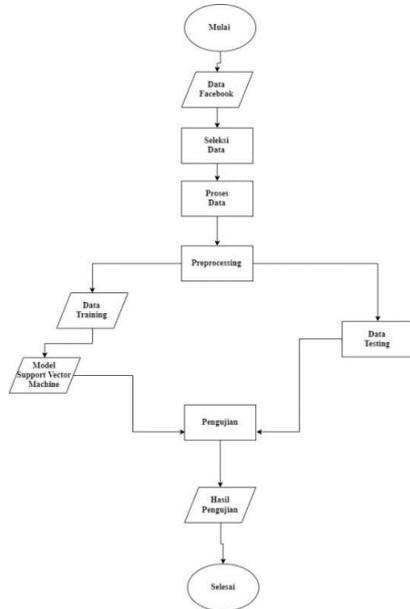
3.3. Rancangan Sistem

Pada tahap ini, dilakukan perancangan sistem untuk menjelaskan cara kerja dari sistem. Rancangan sistem akan digambarkan melalui *flowchart* yang ditunjukkan pada Gambar 4.

Berdasarkan Gambar 4, tahap awal proses adalah pengambilan data dari Facebook maupun Twitter. Selanjutnya terdapat proses seleksi data sesuai dengan kriteria yang ditentukan,

yaitu status media sosial. Kemudian data hasil seleksi diproses menggunakan LIWC bertujuan memperoleh nilai fitur teks pada status Facebook maupun Twitter. Selanjutnya dilakukan *preprocessing* data bertujuan untuk *filtering* untuk memilih atribut/fitur yang diperlukan maupun normalisasi data untuk menghilangkan redundansi data, kemudian data dibagi menjadi dua, yaitu 80% data latih dan 20% data uji. Setelah itu dilakukan

proses pelatihan untuk memperoleh model SVM yang digunakan untuk proses pengujian.



Gambar 4. Alur rancangan sistem

3.4. Implementasi Sistem

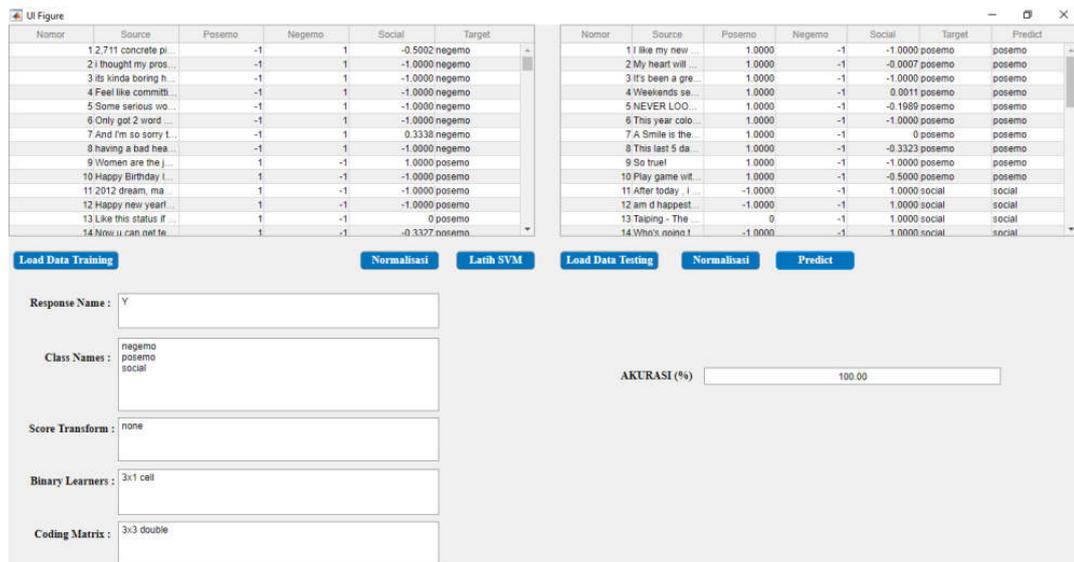
Pada tahap ini, dilakukan pembangunan sistem berdasarkan perancangan yang telah dibuat. Implementasi sistem ini meliputi

penerapan bahasa pemrograman dengan membangun antar muka yang mudah untuk dioperasikan. Gambar 5 diatas merupakan hasil implementasi sistem dalam bahasa pemrograman MATLAB menggunakan antar muka GUI.

3.5. Uji Coba dan Evaluasi Sistem

Pada bagian ini, dilakukan uji coba dan evaluasi sistem terkait prediksi kepribadian pengguna Facebook maupun twitter menggunakan metode Support Vector Machine dilakukan. Evaluasi sistem bertujuan untuk mengamati tingkat akurasi sistem yang ditunjukkan pada Persamaan (10). Dengan tingkat akurasi yang tinggi, maka sistem dapat digunakan sebagai perangkat yang mampu menentukan kepribadian seseorang berdasarkan status media sosial secara efektif.

$$Akurasi = \frac{Jml\ Prediksi\ Benar}{Jml\ Total\ Prediksi} \quad (10)$$



Gambar 5. Program analisa kepribadian menggunakan metode Support Vector Machine

index	screen_name	username	user_id	tweet_id	tweet_url	timestamp	timestamp_epoch	text	text_html	links	hashtags	has_media
0	endowedorid	Owambe Trends Paga...	391382288	113134901574...	/endowedorid...	2019-05-22 23:59:59	1558569599	1. Argentina	<p>Brazil, Para...	[]	[]	True
1	KatiaBarros	Katia Barros	22835817	113134900995...	/KatiaBarros/...	2019-05-22 23:59:58	1558569598	This is how	<p>it is done h...	[]	[]	False
2	carrolbarton6	carrol barton	103452150304...	113134900949...	/carrolbarton...	2019-05-22 23:59:57	1558569597	President	<p>Trump is the...	[]	[]	False
3	OneOpinion8	OneOpinion	189144326624...	113134900728...	/OneOpinion8/...	2019-05-22 23:59:57	1558569597	Ilham Omar	<p>has removed...	[]	[]	False
4	Emilycgn	Emily Best	15472207	113134900672...	/Emilycgn/...	2019-05-22 23:59:57	1558569597	This is rare	<p>for me beca...	[]	['winwithhar...	True
5	WeAreAllVIPs	Elizabeth Post	900651226646...	113134900641...	/WeAreAllVIPs...	2019-05-22 23:59:57	1558569597	It's weird	<p>that a mila...	[]	[]	False
6	Nathan746618	Nathan	101346842655...	113134900246...	/Nathan746618...	2019-05-22 23:59:56	1558569596	We voted to	<p>leaveby over...	[]	[]	False
7	Agape1v53	Boomer	752604131316...	113134900128...	/Agape1v53/...	2019-05-22 23:59:55	1558569595	Love youall	<p>you have don...	[]	[]	False
8	ILLTRILLNEEL	ILL TRILL NEEL	23742003	113134900103...	/ILLTRILLNEEL...	2019-05-22 23:59:55	1558569595	Number1	<p>#Artist and...	[]	['Number1', 'Artist', 'B...	True
9	firecracker1	Gwen	162307520	113134900057...	/firecracker1...	2019-05-22 23:59:55	1558569595	@realDonaldTrump	<p>	[]	[]	False
10	gendng1r1625	Thugardng1...	22018205	113134899647...	/gendng1r162...	2019-05-22 23:59:54	1558569594	#CreepyPorn...	<p>class="Tweet...	['https://...	['CreepyPorn...	False
11	ercraps77	Michelle Cichocki	2372868220	113134899428...	/ercraps77/...	2019-05-22 23:59:54	1558569594	what country	<p>for 30 cents...	[]	[]	False
12	HattCassady93	Hatt Cassidy	35394803	113134899063...	/HattCassady9...	2019-05-22 23:59:53	1558569593	Hollywood	<p>doesn't repr...	[]	[]	False
13	HM66A6en	Aden Pizarro	110318882781...	11313489902...	/HM66A6en/...	2019-05-22 23:59:53	1558569593	Song: Black	<p>Blood Sands...	['https://...	['HeavyMetal...	False
14	SiraSpurg	Sira Spurgeon	2831820887	113134898512...	/SiraSpurg/...	2019-05-22 23:59:52	1558569592	Please don't	<p>forget to me...	[]	[]	False
15	Country_Fan...	Country_Fan...	609127539	113134898387...	/Country_Fan...	2019-05-22 23:59:51	1558569591	Brett Young's	<p>"Catch" Find...	['https://...	['tasteofcount...	False
16	HorganFeldma	Horgan Feldm...	958199613781...	113134898005...	/HorganFeldma...	2019-05-22 23:59:51	1558569591	Graman says	<p>he wants Cen...	[]	[]	False
17	DeeHarcky	Harcky DEE	869664883113...	113134897670...	/DeeHarcky/...	2019-05-22 23:59:50	1558569590	#OFFICIALMas...	<p>class="Tweet...	['https://...	['protecteal...	False
18	JLDixon7	JL Dixon	786889083268...	113134897634...	/JLDixon7/...	2019-05-22 23:59:50	1558569590	Trump stomps	<p>all over our...	['https://...	['EndGerryma...	False
19	places_calli...	/PlaceHatt	2349925688	113134897117...	/places_calli...	2019-05-22 23:59:48	1558569588	...and our	<p>responsibili...	['https://...	['twitter.com/...	False
	Owambe					2019-05-22		1. Argentina	<p>	True

Gambar 6. Hasil pengambilan data status media sosial

4. Hasil dan Pembahasan

Pada tahapan ini, dijelaskan proses pengolahan data hingga diperoleh hasil evaluasi dari prediksi kepribadian seseorang berdasarkan status media sosial menggunakan Support Vector Machine.

Gambar 6 menunjukkan proses *scrapping* untuk memperoleh data status dari Facebook maupun Twitter.

Untuk data status Facebook memiliki total 300 status dan Twitter sebanyak 2067 status, kemudian status tersebut dibagi menjadi 80 persen data latih dan 20 persen data uji. Sehingga, untuk Facebook diperoleh 240 data latih dan 60 data uji, serta untuk Twitter diperoleh 1653 data latih dan 414 data uji.

Data tersebut kemudian diproses menggunakan LIWC untuk memperoleh bobot nilai. Pada tahap awal, setiap status dihitung bobot nilai pada seluruh kategori, yaitu *summary language variables*, *linguistic dimensions*, *other grammar*, dan *psychological process*. Namun, penelitian ini hanya berfokus pada aspek *psychological process*, yang terdiri dari *negative emotion* (emosi negative), *positive emotion* (emosi

positif), dan *social process* (proses sosial). Ketiga aspek tersebut menjadi atribut/fitur set data yang digunakan sebagai nilai parameter ukuran kepribadian seseorang secara psikologis dalam penelitian ini.

Setelah dilakukan proses seleksi terhadap tiga aspek dalam set data pada Facebook maupun Twitter, masing-masing data kemudian dinormalisasi pada nilai range 0 hingga 1 atau [0,1] dan nilai range -1 hingga 1 atau [-1,1]. Tujuan dari normalisasi data ini adalah mengurangi redudansi data. Persamaan (11) dan (12) berikut ini merupakan rumus umum normalisasi pada range [0,1] dan [-1,1]. Untuk normalisasi pada range [0,1]

$$Z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (11)$$

Untuk normalisasi pada range [-1,1]

$$Z_i = 2 \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (12)$$

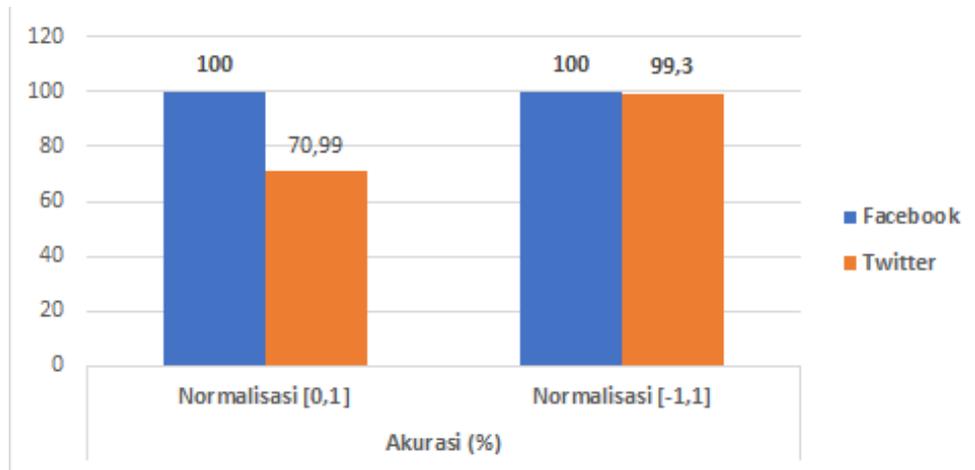
Tujuan perbandingan dua normalisasi tersebut adalah untuk menganalisis normalisasi data yang dapat mempengaruhi performansi maksimal.

Tahapan selanjutnya yaitu proses pelatihan (*training*) menggunakan metode Support Vector Machine dengan *kernel linear*. Pada proses pelatihan ini, diperoleh model Support Vector Machine yang selanjutnya digunakan untuk proses pengujian prediksi terhadap status Facebook maupun Twitter diluar data pelatihan atau dapat dikatakan menggunakan data pengujian. Gambar 7 menunjukkan model Support Vector Machine.

```
svm_latih =
classreg.learning.classif.CompactClassificationECOC
  ResponseName: 'Y'
  ClassNames: {'negemo' 'posemo' 'social'}
  ScoreTransform: 'none'
  BinaryLearners: (3x1 cell)
  CodingMatrix: [3x3 double]

Properties, Methods
```

Gambar 7. Model Support Vector Machine



Gambar 8. Hasil perbandingan evaluasi data uji berdasarkan tingkat akurasi (%)

Seperti halnya data latih, sebelum dilakukan proses uji, data uji perlu dinormalisasi dalam range [0,1] dan [-1,1]. Tabel 2 menunjukkan hasil evaluasi performansi Support Vector Machine dalam memprediksi kepribadian seseorang berdasarkan status di media sosial.

Tabel 2. Hasil Evaluasi Sistem

Data Uji	Akurasi (%)	
	Normalisasi [0,1]	Normalisasi [-1,1]
Facebook	100	100
Twitter	70,99	99,3

Berdasarkan Gambar 8, dapat ditunjukkan bahwa pada data uji

Facebook pada normalisasi *range* [0,1] dan [-1,1] keduanya memiliki akurasi yang sangat optimal 100%. Sementara itu, data uji Twitter memiliki tingkat akurasi optimal 99,3% pada normalisasi *range* [-1,1] dibandingkan dengan normalisasi [0,1] dengan akurasi 70,99%. Ini artinya bahwa pada data uji Twitter pada normalisasi [0,1] masih memiliki nilai redundansi, sehingga sistem tidak dapat memprediksi secara optimal.

5. Penutup

5.1. Kesimpulan

Berdasarkan hasil analisis evaluasi, kesimpulan dari penelitian ini bahwa Support Vector Machine

mampu memprediksi kepribadian berdasarkan status dengan normalisasi data $[0,1]$ untuk Facebook diperoleh akurasi 100% dan untuk Twitter diperoleh akurasi 70,99%. Dengan normalisasi data $[-1,1]$ untuk Facebook diperoleh akurasi 100% dan untuk Twitter diperoleh akurasi 99,3%. Penelitian ini memberikan kontribusi penelitian dalam bidang analisa penggalan data teks.

5.2. Saran

Untuk pengembangan selanjutnya, maka disarankan aspek bahwa penelitian ini dapat dikembangkan untuk menganalisa teks status dalam bahasa Indonesia dan penelitian ini dapat ditambahkan data latih agar dapat meningkatkan akurasi. Penelitian ini dapat dievaluasi performansi waktu komputasinya. Sehingga, mampu memberikan kontribusi dalam bidang optimasi. Penelitian ini dapat dikembangkan dengan menambahkan proses pengambilan data media sosial dan perhitungan nilai fitur teks yang terintegrasi, tanpa memerlukan perangkat lunak kedua. Penelitian ini mampu berkontribusi dalam pengembangan penelitian bidang data mining.

6. Daftar Pustaka

- [1] T. A. D. Permana, F. Sholihin, and F. Hastarita, "Klasifikasi Emosi Teks Berbahasa Indonesia Menggunakan Metode Maximum Entropy," *Konvergensi*, vol. 13, no. 2, pp. 68–75, 2017.
- [2] F. N. Rozi and D. H. Sulistyawati, "Klasifikasi Berita Hoax Pilpres Menggunakan Metode Modified K-Nearest Neighbor Dan Pembobotan Menggunakan Tf-Idf," *Konvergensi*, vol. 15, no. 1, pp. 1–10, 2019.
- [3] D. Markovikj, S. Gievska, M. Kosinski, and D. Stillwell, "Mining facebook data for predictive personality modeling," *AAAI Workshop - Technical Report*, vol. WS-13-01, pp. 23–26, 2013.
- [4] A. Ortigosa, R. M. Carro, and J. I. Quiroga, "Predicting user personality by mining social interactions in Facebook," *Journal of Computer and System Sciences*, vol. 80, no. 1, pp. 57–71, 2014.
- [5] C. Limantara and D. Nababan, "Klasifikasi Kepribadian Menggunakan Algoritma Decision Tree Berdasarkan Ten Item Personality Inventory," *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, vol. 5, no. 1, pp. 8–12, 2019.
- [6] E. Yuliani, E. Utami, and S. Raharjo, "Klasifikasi Kepribadian Pengguna Media Sosial," *Jurnal Informa*, vol. 6, no. 1, pp. 15–19, 2020.
- [7] A. S. R. Sinaga, "Machine Learning Prediksi Karakter Pengguna Hastag (#) Bahasa Generasi Milenial Di Sosial Media," *Indonesian Journal of Applied Informatics*, vol. 4, no. 2, pp. 165–171, 2020.
- [8] D. Lhaksmana, KM; Nhita, Fhira & Anggraini, "Klasifikasi Kepribadian Berdasarkan Status Facebook Menggunakan Metode Backpropagation," *e-Proceeding of Engineering*, vol. 4, no. 3, pp. 5174–5183, 2017.
- [9] Y. B. N. D. Artissa, I. Asror, and S. A. Faraby, "Personality Classification based on Facebook status text using Multinomial Naïve Bayes method," *Journal of Physics: Conference Series*, vol. 1192, no. 1, 2019.
- [10] A. Fikriani, I. Asror, and Y. R.

- Murti, “Klasifikasi Kepribadian Berdasarkan Data Twitter dengan Menggunakan Metode Support Vector Machine,” *e-Proceeding of Engineering*, vol. 6, no. 3, pp. 10436–10450, 2019.
- [11] E. Ronando, M. I. Irawan, and E. Apriliani, “A Hybrid Approach Support Vector Machine (SVM) - Neuro Fuzzy For Fast Data Classification,” *IPTEK Journal of Proceeding Series*, vol. 1, pp. 105–106, 2015.
- [12] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, “The Development and Psychometric Properties of LIWC2015,” Austin, TX: University of Texas at Austin, 2015.
- [13] R. L. Boyd, “Data Analytics in Digital Humanities,” in *Hai-Jew S. (eds) Data Analytics in Digital Humanities. Multimedia Systems and Applications*, Springer, Cham, 2017, pp. 161–189.
- [14] E. Ronando and Sugiono, “Sistem Konversi Ucapan Kata Ke Teks Menggunakan Support Vector Machine: Speech Word Recognition To Text Converter Using ...,” *Jurnal Teknologi dan Terapan Bisnis*, vol. 2, no. 2, pp. 1–8, 2019.
- [15] S. A. Catanese, P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, “Crawling Facebook for social network analysis purposes,” in *WIMS '11: International Conference on Web Intelligence, Mining*, 2011.
- [16] A. Hernandez-Suarez, G. Sanchez-Perez, K. Toscano-Medina, V. Martinez-Hernandez, V. Sanchez, and H. Perez-Meana, “A Web Scraping Methodology for Bypassing Twitter API Restrictions,” *arXiv e-prints*, pp. 1–7, 2018.
- [17] A. Gholamy, V. Kreinovich, and O. Kosheleva, “Why 70/30 or 80/20 Relation Between Training and Testing Sets : A Pedagogical Explanation,” 2018.