

KLASIFIKASI KEPERIBADIAN BERDASARKAN STATUS FACEBOOK MENGUNAKAN METODE *BACKPROPAGATION*

CLASSIFICATION OF PERSONALITY BASED ON FACEBOOK STATUS USING BACKPROPAGATION

Kemas Muslim Lhaksana¹, Fhira Nhita², Duwi Anggraini³

Prodi S1 Ilmu Komputasi, Fakultas Informatika, Universitas Telkom

Kemasmuslim@telkomuniversit.ac.id, duwianggraini@student.telkomuniversity.ac.id

Abstrak

Informasi status facebook dapat dimanfaatkan untuk menggambarkan kepribadian seseorang, yang terdiri dari *social word*, *positive emotions*, dan *negative emotions*. Pada penelitian tugas akhir ini pelamar kerja dianalisa berdasarkan status facebook untuk membentuk suatu model pembelajaran prediksi kepribadian pelamar kerja tersebut. Fitur tersebut didapatkan dari ekstraksi dengan metode *Linguistic Inquiry and Word Count* (LIWC) merupakan program analisis teks yang menghitung sebuah kata-kata dalam kategori psikologis menjadi nilai persentase, seperti *social word*, *positive emotions* dan *negative emotins*. Pada proses pengujian dilakukan dengan menggunakan *backpropagation*, data status yang telah diberi label kelas digunakan untuk pelatihan (*training*). Setelah proses *training* selesai maka hasil *training* diuji dengan parameter ANN menggunakan 2, 4, 6 dan 8 *neuron* pada *hidden layer*. Hasil pengujian kemudian dibandingkan untuk melihat kelebihan dan kekurangan dari *neuron* tersebut. Dari hasil pengujian yang dilakukan dengan *neuron* pada *hidden layer* didapatkan akurasi sebesar 84,00 % dengan nilai *Mean Square Error* (MSE) terkecil sebesar 0.0001 di pengujian 3 dengan data 70:30%.

Kata kunci : Klasifikasi, *Linguistic Inquiry and Word Count* (LIWC), facebook, *backpropagation*

Abstract

Facebook status information can be used to describe someone's personality, consisting of social word, positive emotions, and negative emotions. On the research of this thesis work is analyzed based on the applicants status facebook to establish a predictive learning model of the personality of the applicants. The feature extraction method of *Linguistic Inquiry and Word Count* (LIWC) is a transparent text analysis program that calculates the value of a word in the catagory of psychological. In the process of testing is carried out using the method of *backpropagation*, have classified the State data used for tarining. After the training is completed then the results of the training were tested with parameters ANN using 2,4,6 and 8 of neurons in the hidden layer. The test results are then compared to see the advantages and disadvantages of the hidden neurons. From the results of testing conducted with hidden neurons get accuracy of 84,00 % by value of the Mean Square Error (MSE) of smallest of 0,00001 in test results 3 with 70:30% data.

Keywords: classification, *Linguistic Inquiry and WordCount* (LIWC), facebook, *backpropagation*

1. Pendahuluan

Kepribadian merupakan salah satu metode yang dikenal dalam dunia piskologi untuk menginterpretasi kepribadian seseorang, terutama untuk menemukan hubungan kepribadian dengan lingkungan pekerjaan. Kepribadian seseorang dapat dilihat dari berbagai aspek yaitu kepribadian *openness*, *extraversion* dan lain-lainnya. Pengukuran kepribadian dapat dilakukan dengan menggunakan berbagai metode. Salah satu metodenya adalah berupa kuesioner pertanyaan yang berisi sifat-sifat berbentuk skala dari paling setuju sampai paling yang tidak setuju. Tetapi cara ini memiliki kekurangan karena hasil prediksi dari kuesioner kurang valid[19].

Dalam hal ini akan dibuat model klasifikasi kepribadian berdasarkan status media social yaitu facebook dengan *Linguistic Inquiry Word Count* (LIWC) menggunakan metode *Backpropagation*. Dengan melihat 3 parameter untuk menyeleksi kepribadian yaitu *positive emotion*, *negative emotion* dan *social word*.

Metode *Linguistic Inquiry and Word Count* (LIWC) adalah program analisis teks transparan yang menghitung kata-kata dalam kategori psikologis yang bermakna atau metode analisis kata yang mengubah status tersebut menjadi nilai-nilai. Hasil LIWC menunjukkan kemampuannya untuk mendeteksi makna dalam berbagai macam pengaturan eksperimental, termasuk untuk menunjukkan fokus perhatian, emosionalitas, hubungan sosial, gaya berpikir, dan perbedaan individual[17].

Backpropagation adalah metode klasifikasi yang dapat di implementasikan pada kasus prediksi. *Backpropagation* mampu melakukan proses *learning* yang repressasinya dianalogikan seperti otak manusia. *Backpropagation* termasuk algoritma pembelajaran yang terawasi dengan arsitektur yang sederhana dan mampu mengurangi error[5].

Tujuan dari penelitian tugas akhir ini adalah untuk mengetahui bagaimana tingkat akurasi klasifikasi kepribadian berdasarkan status facebook dengan menggunakan metode *backpropagation* yang perhitungan tingkat akurasinya dinyatakan dalam bentuk *confusion matrix*.

2. Tinjauan Pustaka

2.1 *Linguistic Inquiry and Word Count (LIWC)*

Linguistic Enquiry and Word Count (LIWC) adalah sebuah perangkat lunak penghitung kata yang banyak digunakan untuk analisis teks kuantitatif dalam ilmu sosial. Meskipun LIWC mampu mengukur fitur dalam teks yang memungkinkan klasifikasi teks dan prediksi untuk berbagai hasil perilaku, ini terutama digunakan untuk mengidentifikasi fitur kata yang informatif tentang keadaan psikologis pengarang atau pembicara atau kelompok yang mendasarinya. LIWC pada awalnya dikembangkan untuk menangani masalah analitik konten dalam psikologi eksperimental[10].

Beberapa aplikasi LIWC misalnya, sebagai detektor kebohongan, status atau barometer sosial. Isu utama pendekatan LIWC relatif terhadap pendekatan pengolahan bahasa alami lainnya (NLP) dibahas, bersama dengan beberapa solusi dan rekomendasi. Akhirnya, kami menyajikan cara baru di mana LIWC sedang diterapkan, dan bagaimana hal itu dapat digunakan dalam penelitian selanjutnya. LIWC dapat digunakan sebagai kata benda, kata sifat, dan kata kerja. Penggunaan kata kerja sedang dalam pengembangan[10].

A. Cara kerja *Linguistic Inquiry and Word Count (LIWC)*

Cara kerja *Linguistic Inquiry and Word Count (LIWC)* cukup sederhana. Pada dasarnya, ini membaca teks tertentu dan menghitung persentase kata-kata yang mencerminkan emosi, gaya berpikir, masalah sosial, dan bahkan bagian pembicaraan yang berbeda. Karena LIWC dikembangkan oleh para periset yang memiliki kepentingan dalam psikologi sosial, klinis, kesehatan, dan kognitif, kategori bahasa diciptakan untuk menangkap keadaan sosial dan psikologis masyarakat[18].

2.2 Facebook

Facebook merupakan sarana sosial yang menghubungkan orang-orang dengan teman dan rekan mereka lainnya yang bekerja, belajar, dan hidup di sekitar mereka. Orang-orang menggunakan Facebook untuk menjaga hubungan dengan teman, bertukar foto tanpa batas, mengirim tautan dan video, dan mengetahui lebih jauh tentang orang-orang yang mereka temui. Pengguna dapat membuat profil dilengkapi foto, daftar ketertarikan pribadi, informasi kontak, dan informasi pribadi lain. Pengguna dapat berkomunikasi dengan teman dan pengguna lain melalui pesan pribadi atau umum dan fitur obrolan.

Facebook memiliki sejumlah fitur yang dapat berinteraksi dengan pengguna. Salah satunya adalah dinding, kotak di setiap halaman profil pengguna yang mengizinkan teman mereka mengirimkan pesan kepada pengguna tersebut. Colek, yang memungkinkan pengguna mengirimkan "colekan" virtual satu sama lain (pemberitahuan memberitahu pengguna bahwa mereka telah dicolek). Foto, tempat pengguna dapat mengunggah album dan foto dan status, yang memungkinkan pengguna untuk memberitahukan teman mereka mengenai keberadaan dan tindakan mereka saat itu. Facebook mempunyai mekanisme yang mendukung untuk menyebarkan informasi yang didapat dari hasil tulisan orang lain dengan cara bagikan (*shared*) yang dapat dikirimkan kepada teman atau dikirimkan ke kronologi anda[14].

2.3 Backpropagation

Salah satu dari sekian banyak algoritma pelatihan untuk MLP yang sangat populer adalah *backpropagation* atau Propagasi Balik. Sesuai dengan namanya, algoritma ini melakukan dua tahap perhitungan, yaitu: perhitungan maju untuk menghitung *error* antara keluaran aktual dan target, dan perhitungan mundur yang mempropagasikan balik *error* tersebut untuk memperbaiki bobot-bobot sinpatik pada semua *neuron* yang ada. Berikut ini adalah langkah-langkah detail dari algoritma pelatihan Propagasi Balik untuk MLP dengan satu *hidden layer* (dengan fungsi aktivasi sigmoid biner) [5]:

A. Algoritma pelatihan standar Backpropagation

a) Perhitungan maju

1. Inisialisasi perangkat jaringan, yang terdiri dari:

a. Arsitektur jaringan (*input layer, hidden layer, output layer*)

- b. Nilai ambang (*threshold*)
 - c. Laju pembelajaran atau *Learning rate* (α)
 - d. *Mean Square Error* (*MSE*)
 - e. Bobot-bobot (W)
2. Inisialisasi semua bobot dengan bilangan acak kecil secara *random*.
 3. Hitung semua keluaran diunit tersembunyi atau untuk mendapatkan nilai hidden layer Z_j ($j=1,2,3,\dots,p$)

$$Z_{netj} = b + \sum_{i=1}^n x_i z_{ij} \quad (2.1)$$

$$Z_j = F(Z_{netj}) = \frac{1}{1 + e^{-Z_{netj}}} \quad (2.2)$$

Dengan :

Z_{netj} : keluaran nilai *hidden layer*

z_{ij} : bobot antar *input layer* dan *hidden layer*

Z_j : cek nilai fungsi aktivasi (sigmoid)

4. Hasil keluaran dari *hidden layer* dipakai untuk mendapatkan keluaran *output layer* y_k ($k = 1,2,3,\dots,m$) menggunakan persamaan :

$$Y_{net} = b + Z_{net1}(y_1) + Z_{net2}(y_2) \quad (2.3)$$

Dengan :

Z_{netj} : Keluaran nilai *hidden layer*

y : bobot sinaptik antar *output layer* dan *hidden layer*

b : nilai *bias* simpul *hidden*

Y_{net} : keluaran *output layer*

$$y_k = F(y_{netk}) = \frac{1}{1 + e^{-Y_{net}}} \quad (2.4)$$

Dimana y_k untuk nilai aktivasi yang sudah didapatkan.

5. Membandingkan nilai target
Keluaran dari hasil jaringan ini *output* dibandingkan dengan nilai target. Selisih antara nilai target dengan keluaran jaringan adalah nilai *error* (E) yang dapat dirumuskan dengan:

$$E = T - Y_{net} \quad (2.5)$$

Dengan :

T : matrik target

Y_{net} : Keluaran *output layer*

Kemudian nilai rata-rata kuadrat *error* atau *MSE* dinyatakan oleh persamaan :

$$MSE = \frac{\sum E^2}{N} \quad (2.6)$$

Dimana N adalah jumlah pola masukan.

- b) Perhitungan Mundur

MSE yang diperoleh dipakai sebagai parameter dalam pelatihan. Pelatihan akan selesai jika *MSE* yang diperoleh sudah dapat diterima. *Error* tersebut dipropagasikan balik untuk memperbaiki bobot sinaptik dari semua neuron pada *hidden layer* dan *output layer*. Perhitungan perbaikan bobot diberikan pada persamaan berikut:

1. Cari faktor kesalahan (δ) di unit keluaran.

$$\delta = (tk - yk) \cdot yk \cdot (1 - y) \quad (2.7)$$

Dimana :

δ : unit kesalahan yang digunakan dalam perubahan bobot

tk : pola target

yk : nilai aktivasi yang sudah didapatkan

2. Hitung suku perubahan bobot w_{jk} yang akan digunakan untuk merubah bobot dengan laju pelatihan.

$$\Delta w_{kj} = \alpha \cdot \delta \cdot z_{ij} \quad (2.8)$$

Dimana :

z_{ij} : nilai aktivasi dari *input* ke *hidden*

hitung perubahan *bias*

$$\Delta b = \alpha \cdot \delta k \quad (2.9)$$

Dimana :

Δb : nilai *bias*

3. Hitung koreksi kesalahan bobot dan *bias* antara lapisan *input* dilapisan tersembunyi

$$\delta_{netj} = \sum_{k=1}^m \delta_k w_{kj} \quad (2.10)$$

Hitung aktivasi

$$\delta_i = \delta'_{netj} (\delta_{netj}) = \delta_{netj} z_j (1 - z_j) \quad (2.11)$$

4. Hitung perubahan bobot V_{ji} (nilai antara lapisan *input* dilapisan tersembunyi)

$$\Delta V_{ji} = \alpha \cdot \delta_j \cdot X_i \quad (2.12)$$

Hitung perubahan *bias*

$$\Delta V_{jk} = \alpha \cdot \delta_j \quad (2.13)$$

5. *Update* bobot dan *bias*

Hitung perubahan bobot *hidden - output* yang akan menghasilkan bobot dan *bias* baru :

$$W_{kj}(\text{baru}) = W_{jk}(\text{lama}) + \Delta W_{jk} \quad (2.14)$$

Kemudian untuk setiap unit tersembunyi mulai dari ke-1 sampai n dilakukan pengupdetan bobot dan *bias*.

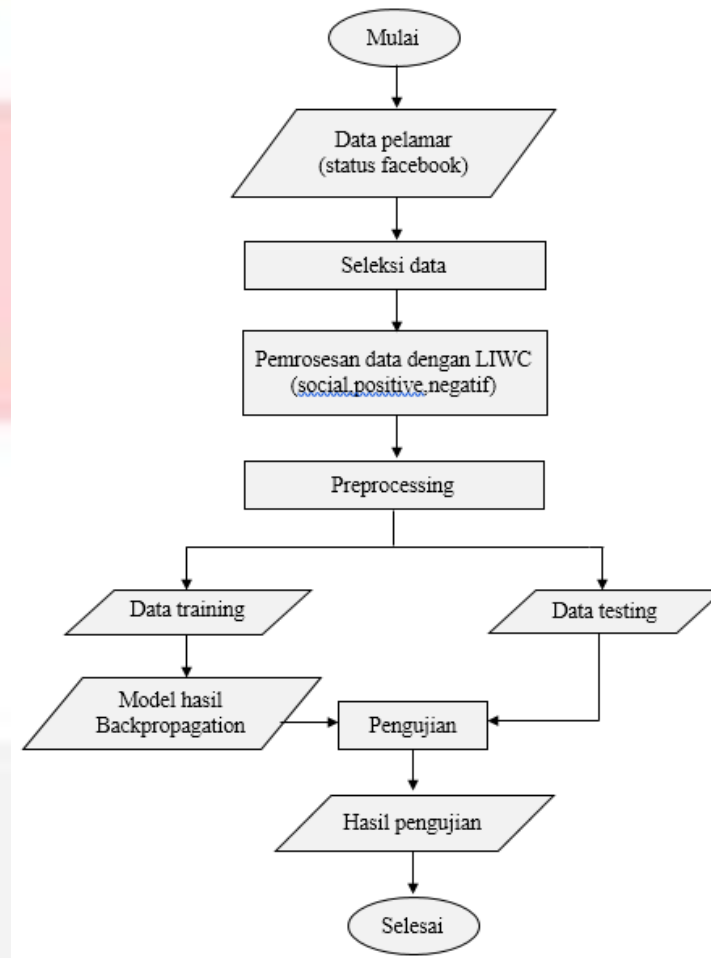
$$V_{ji}(\text{baru}) = V_{ji}(\text{lama}) + \Delta V_{ji} \quad (2.15)$$

6. Pelatihan dihentikan ketika (max epoch) dan telah tercapai MSE (*Mean Square Error*) yang diinginkan.

3. Metodologi Penelitian

3.1. Gambaran umum

Gambaran umum merupakan alur kerja dari sistem yang membantu untuk membuat rancangan dari proses penelitian dari awal sampai akhir. Berikut merupakan *flowchart* dari rancangan sistem mengenai klasifikasi kepribadian berdasarkan informasi pada akun media sosialnya seperti status facebook. Adapun tujuan klasifikasi ini adalah untuk mempermudah mengenali, membandingkan dan mempelajari. Membandingkan berarti mencari persamaan dan perbedaan sifat atau ciri pada suatu hal.



Gambar 3.1 flowchart Alur Perancangan Sistem

Berdasarkan gambar di 3.1 di atas adalah flowchart alur dari sistem yang dibuat. Yang memiliki beberapa proses. Dimana proses pertama adalah menyiapkan data status facebook pada setiap akun. Kemudian data harus melalui proses preprosesing diantaranya adalah seleksi data. Dimana data harus diseleksi untuk diterjemahkan ke dalam bahasa inggris dan menghilangkan singkatan pada status, dan isi konten status itu sendiri sebagai kelas yang akan digunakan dalam pengujian. Kemudian data harus dibagi menjadi dua bagian yaitu data training dan data testing. Pada pengujian backpropagation memerlukan data training sebagai proses belajar untuk membentuk model pola. Setelah backpropagation melakukan proses pembelajaran, maka data testing akan digunakan sebagai data masukan atau inputan. Data testing akan di proses menggunakan propagasi maju. Dalam tahap akhir sistem akan mencatat perormansi proses pembelajaran (learning) dan percobaan (testing).

3.2. Data

Pengumpulan data pelamar yaitu berupa data status facebook. Di mana pada penelitian ini hanya menggunakan 100 orang pelamar yang akan di ambil maksimal 5 status dari setiap akun. Jadi total data ada 468 data status dari seluruh pelamar.

Tabel 3.1 Spesifikasi Data Akun

Data	Jumlah attribute	Jumlah record	Kelas
100 pelamar	3	468	2

Pada proses pengujian, menggunakan tiga porposi pembagian data training dan testing dengan pembagian sebesar 50% training dengan 50% testing untuk pengujian 1, 60% training dengan 40% testing untuk pengujian 2 dan 70% training dengan 30% testing untuk pengujian 3.

Tabel 3.2 Skenario 1 dan 2

Pengujian ke-	jumlah data
1	50 % (<i>training</i>) 50 % (<i>testing</i>)
2	60 % (<i>training</i>) 40 % (<i>testing</i>)
3	70 % (<i>training</i>) 30 % (<i>testing</i>)

4. Hasil Dan Analisis

Berdasarkan hasil uji yang telah dilakukan, didapatkan nilai dari setiap *neuron hidden layer* dengan melihat hasil perhitungan pada analisis status. Adapun nilai didapatkan dari pengujian 1, pengujian 2 dan pengujian 3.

Tabel 4.1 Hasil Pengujian 1

<i>Neuron pada hidden layer</i>				
	<i>2 neuron</i>	<i>4 neuron</i>	<i>6 neuron</i>	<i>8 neuron</i>
Hasil <i>precision</i>	36.36%	38,89 %	28,20 %	31.82%
Hasil <i>recall</i>	50.00%	63,64%	36,36 %	53.85%
Hasil <i>f-measure</i>	42.11%	48,18 %	32,00 %	40.00%
Akurasi	63.33%	63,41 %	58,34 %	58.00%

Berdasarkan pengujian diatas dengan menggunakan metode *Backpropagation* yang dilakukan menggunakan 50 % data *training* dan 50 % data *testing* menunjukkan hasil akurasi prediksi dari proses *testing* tertinggi sebesar 63,41 % di 4 *neuron* pada *hidden layer* dengan hasil *presicion* sebesar 38,89 %, *recall* sebesar 63,64 % dan *f-measure* sebesar 48,18 %. Hal ini dapat disimpulkan bahwa di 4 *neuron* tingkat kebenaran sistem dalam mengklasifikasikan kelas 0 dan kelas 1 dengan optimal.

Tabel 4.2 Hasil Pengujian 2

<i>Neuron pada hidden layer</i>				
	<i>2 neuron</i>	<i>4 neuron</i>	<i>6 neuron</i>	<i>8 neuron</i>
Hasil <i>precision</i>	66,67 %	54.55%	52,94 %	47.06%
Hasil <i>recall</i>	64,55 %	46.15%	81,82 %	61.54%
Hasil <i>f-measure</i>	60,00 %	50.00%	64,29 %	53.33%
Akurasi	80,00 %	76.00%	75,60%	72.00%

Berdasarkan hasil uji diatas, pengujian untuk data ke-2 dengan jumlah data 60 % data training dan 40 % data testing yang diujikan menggunakan Backpropagation menunjukan hasil akurasi testing yang tertinggi sebesar dengan 80,00 % di 2 neuron pada hidden layer dengan hasil presicion sebesar 66,67 %, recall sebesar 81,82 % dan f-measure sebesar 60,00%.

Tabel 4.3 Hasil Pengujian 3

<i>Neuron pada hidden layer</i>				
	<i>2 neuron</i>	<i>4 neuron</i>	<i>6 neuron</i>	<i>8 neuron</i>
Hasil <i>precision</i>	62.50%	50.00%	77.78 %	45.45%
Hasil <i>recall</i>	62.50%	75.00%	53.85%	62.50%
Hasil <i>f-measure</i>	62.50%	60.00%	63.64%	52.63%
Akurasi <i>testing</i>	80.00%	73.33%	84.00%	70.00%

Berdasarkan pengujian diatas dengan menggunakan metode Backpropagation yang dilakukan menggunakan 70 % data training dan 30 % data testing menunjukan hasil akurasi prediksi dari proses testing tertinggi sebesar 84,00 % di 6 neuron pada hidden layer dengan hasil presicion sebesar 77,78 % di 6 neuron , recall sebesar 75,00 % di 4 neuron dan f-measure sebesar 63,64 % di 6 neuron. Hal ini dapat disimpulkan dengan hasil precision menunjukan sistem dapat mengklasifikasikan kelas 0 dan kelas 1 dengan nilai terbesar terdapat di 6 neuron dengan optimal, hasil recall tingkat kebenaran sistem dalam mengklasifikasikan satu kelas (kelas 1) dengan benar.

5. Analisis

Berdasarkan hasil uji yang dilakukan dapat dilihat bahwa pengujian pada setiap neuron hidden memiliki hasil yang berbeda-beda, akan tetapi dengan menggunakan data training lebih banyak dapat menghasilkan nilai akurasi yang baik disetiap pengujian. Dan pada penelitian penggunaan neuron hidden layer tidak bisa menentukan neuron mana yang paling bagus menentukan nilai, karena neuron itu trial dan error, tidak ada ilmu pasti untuk menentukan dengan neuron sekian hasilnya akan baik.

Untuk nilai *precision* merupakan tingkat kebenaran sistem dalam mengklasifikasikan kelas 0 dan kelas 1, jadi kalau di neuron hidden sekian nilai akurasi jelek berarti neuron tersebut tidak mampu mengklasifikasikan 2 kelas dengan optimal. Terkadang jika terlalu banyak neuron data bisa semakin kompleks, jadi kalau tambah kompleks sistem akan semakin susah mengklasifikasikan kedua kelas tersebut. Sedangkan untuk nilai *recall* merupakan tingkat kebenaran sistem dalam mengklasifikasikan kelas, katakan status bernilai kelas 1 dengan keseluruhan data kelas 1.

Untuk nilai *F-measure* merupakan nilai akurasi dari setiap kelas, perhitungan *F-measure* sendiri kombinasi dari nilai *precision* dan *recall*. Dalam penelitian ini *F-measure* digunakan karena kedua data kelas tidak seimbang dengan jumlah datanya dan fokusnya ke prediksi (kepribadian seseorang). Dan untuk nilai akurasi merupakan nilai akurasi dari data testing yang menilai secara keseluruhan.

Hidden layer yang digunakan dalam penelitian ini hanya menggunakan 2 *hidden layer*. Untuk pemilihan hidden layer tergantung dari nilai yang dikeluarkan oleh sistem, terkadang dengan menggunakan 1 hidden layer saja sudah cukup baik. Tetapi, untuk penelitian ini menggunakan 2 hidden layer, karena kalau dengan 1 hidden layer sistem belum bisa mengklasifikasikan secara optimun.

Dalam hal ini nilai linguistik dari LIWC juga digunakan. Dimana pada penelitian ini LIWC hanya menghitung nilai persentase dari setiap kata-kata status dengan kategori yang ditentukan (*posemo, negemo, sosical word*) yang dibagi dengan jumlah panjang kata status.

6. Kesimpulan

Berdasarkan hasil pengujian yang dilakukan menggunakan metode *backpropagation* menghasilkan performansi sistem yang baik dan mendapatkan nilai *error* dengan min kesalahan 0,00001 % pada pengujian 2 dan 3.

Hasil pengujian yang telah dilakukan hasil performansi sistem dengan akurasi tertinggi 84,00 % didapatkan pada proporsi data 70:30 % data yang diujikan.

Pada penelitian penggunaan *neuron hidden layer* tidak bisa menentukan neuron mana yang paling bagus menentukan nilai, karena neuron itu trial dan error, tidak ada ilmu pasti untuk menentukan dengan neuron sekian hasilnya akan baik.

DAFTAR PUSTAKA

- [1] Evanthia Faliagka, Kostas Ramantas, Athanasios Tsakalidis.2012. Application of Machine Learning Algorithms to an online Recruitment System. Computer Engineering and Informatics Department University of Patras Patras, Greece.
- [2] Suyanto. 2011. Artificial Intelligence: Searching, Reasoning, Planning and Learning. Bandung: Informatika
- [3] Jun Gu, Paul w. Purdom, John Franco, and Benjamin.W.Wah.,1991, Algorithms for the satisfiability (SAT) problem : A Survey
- [4] Tsang, Edward.2014.Foundations of Constraint Satisfaction: The Classic Text. BoD–Books on Demand.
- [5] Suyanto., (2008). *Soft Computing*. Bandung :Teknik informatika
- [6] Search in Artificial Intelligence (Springer-Verlag, 1988) and Parallel Algorithms for Machine Intelligence and Vision (Springer- Verlag, 1990).
- [7] Nuryanta, Nanang.2008."Pengelolaan Sumber Daya Manusia (Tinjauan Aspek Rekrutmen dan Seleksi)." Jurnal Pendidikan Islam 1.1.
- [8] Cindy K. Chung., 2012, Linguistic Inquiry and Word Count (LIWC): Pronounced “Luke,” and Other Useful Facts Chapter 12, The University of Texas at Austin, USA.
- [9] Pennebaker, James W., Roger J. Booth, and Martha E. Francis. "Linguistic inquiry and word count: LIWC [Computer software]." *Austin, TX: liwc. net* (2007).
- [10] E. Faliagka, L. Kozanidis, S. Stamou, A. Tsakalidis and G. Tzimas, “Personality Mining System for Automated Applicant Ranking in Online Recruitment Systems,”Proc. Of ICWE 2011, Springer-Verlag, Berlin, Heidelberg, June. 2011, pp. 379-382.
- [11] J.A. Gill, S. Nowson and J. Oberlander, “What are they blogging about? Personality, topic, and motivation in blogs”, Proc. of AAAI ICWSM. 2009.
- [12] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. Witten, “The WEKA data mining software: an update,” SIGKDD Explorer, News, vol. 11, 2009, pp. 10-18.
- [13] Guofang Li and Xiaopeng Ni (2014). Computational Linguistics: Concepts, Methodologies, Tools, and Applications (pp. 1374-1390)
- [14] Mezrich, ben (2009). *The Accidental Billionaires: The Founding of Facebook, A Tale of Sex, Money, Genius, and Betrayal.*, amerika serikat.
- [15] Pant. Gautam,. Srinivasan. Padmini., Menczer, Flippo. “ *Crawling the Web*”. Department of Management Sciences. Universit of Iowa, Iowa city, USA.
- [16] F. Mairesse, M.A. Walker, M.R. Mehl and R.K. Moore,“Using linguistic cues for the automatic recognition of personality in conversation and text,” Journal of Artificial Intelligence Research, vol. 30, 2007, pp. 457-500
- [17] J.W. Pennebaker and L. King, “Linguistic Styles: Language Use as an Individual Difference,” Journal of Personality and Social Psychology, vol. 77, 1999, pp. 1296–1312.

- [18] Yla R. Tausczik and James W. Pennebaker.2009. the psychological of word: LIWC and computerized text analysis methods, *journal of Language and Social Psychology* 2010 29: 24 originally published online 8 december 2009.
- [19] Agnes, theresia dan leylia.2015. prediksi kepribadian big 5 pengguna Twitter dengan support vector regression, *jurnal cybermatika* vol.3 no.1.