

KLASIFIKASI PENGGUNA MEDIA SOSIAL TWITTER DALAM PERSEBARAN HOAX MENGGUNAKAN METODE BACKPROPAGATION

CLASSIFICATION OF USERS SOCIAL MEDIA TWITTER IN THE HOAX SPREAD USING BACKPROPAGATION METHOD

Kemas Muslim Lhaksana¹, Fhira Nhita², Ageng budhiarto³

Prodi S1 Ilmu Komputasi, Fakultas Informatika, Universitas Telkom

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

Abstrak

Penggunaan teknologi akan selalu diikuti dengan hal negatif dan positif pada penggunaannya termasuk seperti pada media sosial twitter, facebook, dan instagram. Banyak kasus yang terjadi seperti penculikan, pencemaran nama baik, penipuan hingga penyebaran suatu paham yang sesat. *Hoax* merupakan salah satu hal negatif yang sering terjadi didalam media sosial, berita dalam *hoax* masih diragukan kebenaran atau faktanya. Penggunaan metode backpropagation pada prediksi dan klasifikasi data dapat digunakan untuk memprediksi kemungkinan seorang pengguna akun twitter menyebarkan suatu berita *hoax* berdasarkan perilaku pengguna tersebut. Pengujian data dilakukan berdasarkan isi konten tweets dan perilaku pengguna. Data set disusun berdasarkan attribute yang digunakan seperti jumlah followers, following, jumlah tweets dan keaktifan pengguna yang berdampak pada diterimanya dan tersebarnya suatu berita. Data set tersebut digunakan untuk pelatihan (training) pada backpropagation menggunakan algoritma gradient descent backpropagation dan lavenberg-marquard backpropagation. Setelah proses training selesai maka hasil training diuji untuk mengenal 4 jenis pola masukan. Hasil pengujian kemudian dibandingkan untuk melihat kelebihan dan kekurangan kedua algoritma backpropagation tersebut. Pada proses pengujian yang dilakukan dengan metode lavenberg-marquard backpropagation mendapatkan akurasi rata-rata sebesar 72.19% dengan MSE (0.1996) terendah dibandingkan dengan backpropagation gradien descent.

Kata kunci : *hoax*, twitter, *Backpropagation*, Neural Network, Media Sosial, *lavenberg-marquard*, *backpropagation*, *gradien descent backpropagation*.

Abstract

The use of technology will always be followed by negative and positive on its use including such as social media twitter, facebook, and instagram. Many cases such as kidnapping, defamation, fraud to the spread of a misguided understanding. *Hoax* is one of the negative things that often occur in social media, news in *hoax* still doubt the truth or fact. The use of backpropagation methods on the prediction and classification of data can be used to predict the possibility of a twitter account user spreading a *hoax* news based on the user's behavior. Testing data is done based on the content of tweets and user behavior. Data sets are arranged based on attributes used such as the number of followers, following, number of tweets and the activity of the user that impact on the receipt and spread of a news. The data sets are used for training on backpropagation using gradient descent backpropagation and lavenberg-marquard backpropagation algorithms. After the training process is complete the training results are tested to recognize 4 types of input patterns. The test results are then compared to see the advantages and disadvantages of both backpropagation algorithms. In the process of testing conducted by the method of lavenberg-marquard backpropagation obtained an average accuracy of 72.19% with the lowest MSE (0.1996) compared with backpropagation gradient descent.

Keywords: *hoax*, twitter, *backpropagation*, Neural Network. Social media, *lavenberg-marquard*, *gadien descent backpropagation*.

1. Pendahuluan

Pada situs jejaring sosial twitter terdapat banyak hal positif dan negatif pada penggunaannya. Twitter selain memudahkan orang untuk berbagi informasi. Juga membawa dampak buruk, seperti penyalahgunaan informasi (pencemaran nama baik, *hoax*, gosip, kampanye hitam, penipuan, pornografi, dan yang lainnya).

Hoax merupakan berita yang tersebar tanpa diketahui kebenaran atau faktanya[2]. Umumnya *hoax* muncul ketika seseorang berusaha untuk menipu atau mengakali seseorang untuk mempercayai sesuatu. *Hoax* tersebar melalui mekanisme pembicaraan antar orang atau tulisan yang ada di media sosial, sehingga menyebabkan *hoax* tersebar dengan cepat. Dengan demikian untuk dapat mengenali ciri-ciri pengguna situs jejaring sosial twitter yang berpotensi menyebarkan berita *hoax*, maka dibutuhkan metode yang dapat mempelajari dan membaca ciri-ciri tersebut dan dalam tugas akhir ini akan menggunakan metode jaringan syaraf tiruan untuk mengklasifikasikan pengguna yang berpotensi menyebarkan suatu berita *hoax* berdasarkan perilaku pengguna didalam twitter.

Metode *backpropagation* merupakan salah satu program komputasi untuk penerapan neural network atau jaringan syaraf tiruan yang sering digunakan untuk memecahkan masalah non-linier serta network multilayer [10]. *Backpropagation* menggunakan pelatihan terbimbing atau terstruktur (train neural network) dan dalam pengaturan jumlah lapisan (layer) mudah dilakukan sehingga mudah diterapkan dalam berbagai masalah. *Backpropagation* mampu melakukan proses *learning* (belajar) seperti halnya otak manusia yang mampu menyimpan semua pengetahuan yang sudah di pelajarnya dan menghasilkan tingkat *error* yang terkecil.

Dalam penelitian ini penulis menggunakan metode Jaringan Syaraf Tiruan (JST) dengan algoritma *Backpropagation* (BP) yang digunakan untuk proses *learning*. Penelitian menggunakan metode neural network sebelumnya pernah dilakukan untuk branch prediction [14] dan kadar polutan[15]. Metode *backpropagation* dianggap metode yang cocok untuk mendeteksi perilaku pengguna yang berpotensi menyebarkan *hoax* yang dapat menghasilkan tingkat error terkecil dan tingkat akurasi dinyatakan dalam bentuk confusion matrix dengan menggunakan metode evaluasi F1-measure.

2. Tinjauan Pustaka

2.1 Hoax

Hoax seringkali muncul dalam pemberitaan dalam suatu media, media cetak maupun media sosial. *Hoax* sendiri berasal dari bahasa inggris yang artinya tipuan, menipu, berita bohong, berita palsu dan kabar burung. Menurut Robert Nares, kata *hoax* muncul sejak abad 18 yaitu "hocus" yang berarti permainan sulap. Sehingga *hoax* dapat diartikan sebagai ketidakbenaran suatu informasi. Pada umumnya *hoax* tersebar diawali dari perbincangan suatu berita dan menyimpulkannya tanpa ada pengetahuan berita tersebut sebelumnya. Biasanya seorang yang menyebarkan berita *hoax* secara sadar dan melakukan suatu kebohongan terhadap berita yang disebar luaskan. Hal ini bertujuan menggiring opini dan kemudian membentuk suatu persepsi yang tidak berdasar dari suatu informasi.

2.2 Twitter

Twitter adalah sebuah situs web yang dimiliki dan dioperasikan oleh Twitter Inc, yang menawarkan jaringan sosial berupa mikroblog sehingga memungkinkan penggunanya untuk mengirim dan membaca pesan Tweets [11]. Tweet adalah penulisan teks yang memiliki 140 karakter yang berada di halaman publik twitter, pengguna bisa membatasi pengiriman tweet atau pesan yang hanya dapat dilihat oleh teman-teman atau pengikut (Followers). Pengguna dapat memeriksa orang-orang yang membatalkan pertemanan (Unfollowing) melalui layanan yang ada di dalam twitter. Pengguna dapat menulis pesan dan mengelompokkannya berdasarkan topik atau jenis pesan dengan menggunakan tagar "#" (hashtag). Pengguna juga dapat menghubungkan pesan Tweet mereka dengan orang lain menggunakan tanda "@" yang diikuti dengan nama pengguna yang akan dihubungkan untuk mengirim atau membalas kiriman pada pengguna lain. Twitter mempunyai mekanisme yang mendukung untuk menyebarkan informasi yang didapat dari hasil tulisan orang lain yaitu retweet. Tweet dibatasi dengan 140 karakter.

2.3 Backpropagation

Backpropagation merupakan salah satu program komputasi untuk penerapan neural network atau jaringan syaraf yang sering digunakan untuk memecahkan masalah non-linier serta network multilayer [10]. *Backpropagation* menggunakan pelatihan terbimbing atau terstruktur (train neural network) dan dalam pengaturan jumlah lapisan (layer) mudah dilakukan sehingga mudah diterapkan dalam berbagai masalah. *Backpropagation* merupakan sistem jaringan yang dapat menghitung tingkat kesalahan dari hasil keluarannya dan akan menampilkan hasil kesalahan (error) yang terkecil.

Algoritma Pelatihan *Backpropagation*:

Berikut algoritma *backpropagation* :

1. Inisialisasi perangkat jaringan, perangkat jaringan terdiri dari :
 - a. Arsitektur jaringan (*input layer, hidden layer, output layer*)
 - b. Nilai ambang (*threshold*)
 - c. Laju pembelajaran atau *Learning rate* (*lr*)
 - d. *Mean Square Error* (*MSE*)
 - e. Bobot-bobot sinaptik (*W*)
2. Inisialisasi bobot
Memasukan nilai bobot dan *bias* untuk tiap simpul secara *random*.
3. Pelatihan jaringan maju
Menghitung tingkat aktivasi
 1. Tingkat aktivasi dari simpul sumber atau *input* tidak perlu dihitung
 2. Setelah pemberian bobot pada inisialisasi awal, kemudian hitung keluaran dari *layer* tersembunyi. Kemudian perhitungan dilakukan menggunakan fungsi aktivasi sigmoid dengan $\alpha = 1$, yang dinotasikan sebagai berikut:

$$\alpha_1 = \frac{1}{1 + e^{-(W_{ij} + P_i + B_1)}} \quad (2.2)$$

Dengan :

P_i : nilai input (matriks pola masukan)

W_{ij} : bobot sinaptik antar input layer dan hidden layer

b_1 : nilai bias simpul hidden

α_1 : keluaran nilai hidden

3. Menghitung tingkat aktivasi dari simpul *hidden* dan *output* menggunakan rumus:

- Untuk simpul *hidden*

$$A_1 = \sum W_{ij} \cdot P_i + b_j \quad (2.3)$$

- Untuk simpul *output*

$$w_k = \sum W_{jk} \cdot A_1 + b_k \quad (2.4)$$

Dengan:

P_i : nilai *input*

W_{ij} : bobot ke simpul hidden

W_{jk} : bobot ke simpul output

b_j : nilai bias simpul hidden

b_k : nilai bias simpul output

A_1 : keluaran nilai hidden

4. 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:

$$E = T - A_2 \quad (2.5)$$

Dengan :

T : matriks target

A_2 : 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.

4. Pelatihan jaringan Mundur

MSE yang diperoleh dipakai sebagai parameter dalam pelatihan. *Error* tersebut dipropagasikan balik untuk memperbaiki bobot-bobot sinaptik dari semua *neuron* pada *hidden layer* dan *output layer*. Perhitungan perbaikan bobot diberikan pada persamaan-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.

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

Dimana :

z_i : 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

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

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* – *ouput* yang akan menghasilkan bobot dan *bias* baru :

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

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

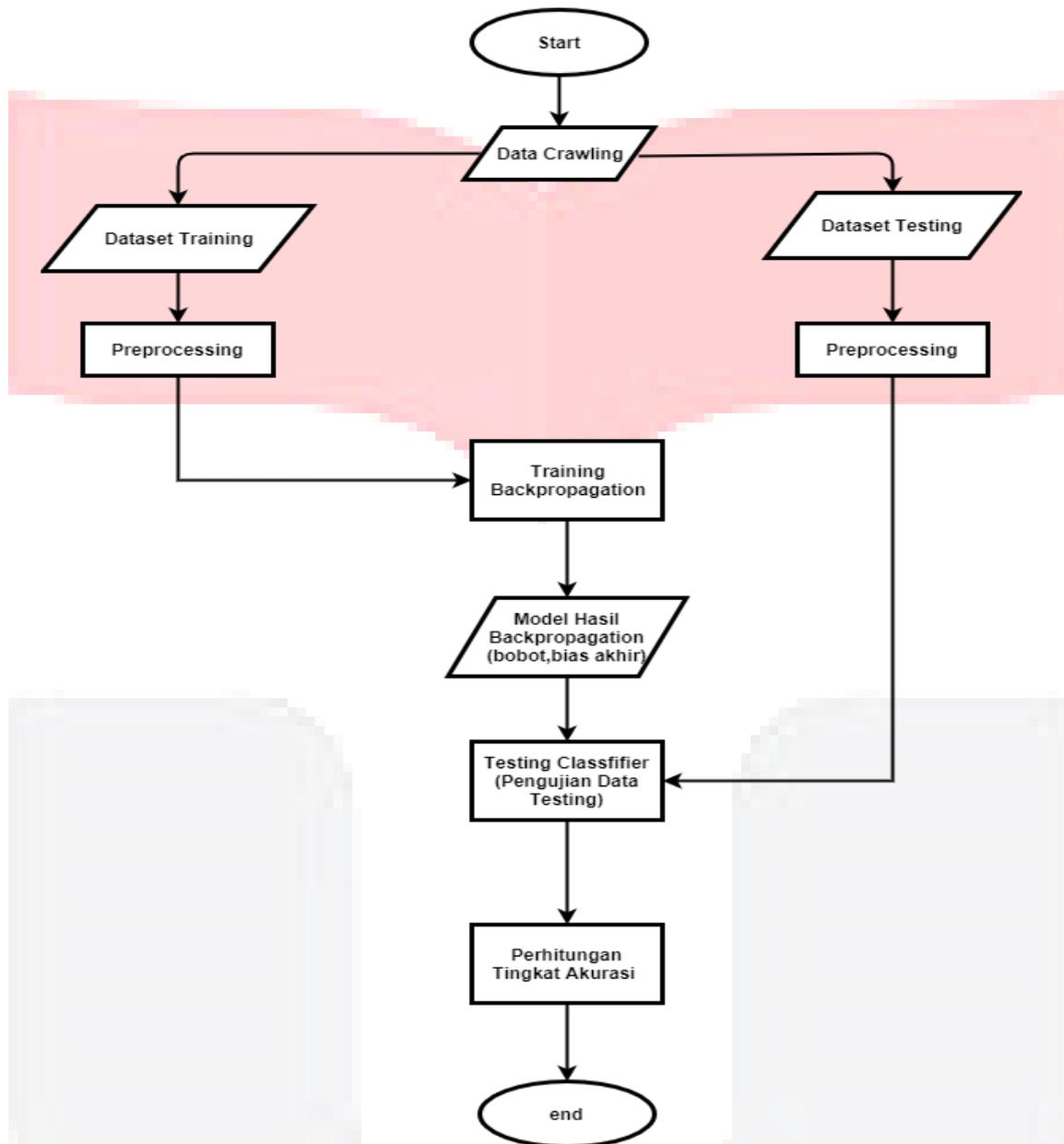
$$V_{ji}(\text{baru}) = V_{ji}(\text{lama}) + \Delta V_{ji} \quad (2.15)$$

6. Pelatihan dihentikan ketika (epoch \geq maxepoch) 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 pengguna media sosial twitter dalam penyebaran *hoax*. Adapun tujuan klasifikasi 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 twitter dengan teknik crawling. Kemudian data harus melalui proses preprosesing diantaranya adalah seleksi data. Dimana data hasil crawling harus diseleksi berdasarkan attribute, dan isi konten tweet 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

Data yang digunakan adalah data akun yang langsung diambil dari media sosial twitter dengan kata kunci #RIPbritney yang berkaitan dengan *hoax*. Pada pengerjaan tugas akhir ini difokuskan terhadap perilaku pengguna dengan *attribute* sebanyak 5, dan menggunakan isi konten tweet untuk mengelaskan data sesuai kriteria.

Tabel 3.1 Spesifikasi Data Akun

Data	Jumlah Data	Jumlah <i>attribute</i>	Jumlah Record	Jumlah kelas
#RIPBritney	1315 akun	5	369(berpotensi) 946(tidak berpotensi)	2

Pada proses pengujian, menggunakan dua porposi pembagian data training dan testing dengan pembagian sebesar 70% training dengan 30% testing untuk skenario 1 dan 50% training dengan 50% testing untuk skenario 2.

Tabel 3.2 Skenario 1 dan 2

Pengujian ke-	Jumlah Data (akun)
1	921(<i>training</i>)394(<i>testing</i>)
2	667(<i>training</i>)658(<i>testing</i>)

Proporsi data yang digunakan untuk proses pengujian pertama sebanyak 70% untuk data training dan 30% untuk data Testing. Pada pengujian kedua digunakan data sebanyak 50% training dan 50% testing.

4. Hasil Dan Analisis

Proses pengujian menggunakan 2 skenario klasifikasi dengan 3 nilai belajar (*learning rate*) yang diujikan untuk membandingkan hasil nilai error terkecil yg didapatkan dari 3 nilai belajar tersebut. Pengujian dilakukan menggunakan metode backpropagation setiap skema pengujian dilakukan 3 kali percobaan untuk membangun model agar mendapatkan hasil arsitektur yang baik. Kemudian dibandingkan dengan hasil nilai akurasi tiap percobaan.

Tabel 4.1 Hasil Pengujian Skenario 1

Klasifikasi	<i>Learning rate</i>	<i>Error rate</i>	epoch	MSE	Akurasi
<i>Backpropagation Lavenberg-Marquadt</i>	0,1	27,66%	8	0,2036	72,34%
	0,5	27,62%	9	0,2024	72,08%
	0,8	27,92%	10	0,1967	72,08%
<i>Backpropagation gradien descent</i>	0,1	28,17%	1000	0,2017	71,83%
	0,5	28,17%	1000	0,2015	71,83%
	0,8	28,17%	1000	0,2016	71,83%

Berdasarkan Pengujian diatas menggunakan metode bacpropagation lavenberg-marquadt dan backpropagation gradien descent yang dilakukan menggunakan 921 data training dan 394 data testing menunjukkan hasil pada metode *bacpropagation lavenberg-marquadt* memiliki performansi diatas 72% dengan MSE lebih kecil dari *backpropagation gradien descent* dengan nilai 0,1967.

Tabel 4.2 Hasil Pengujian Skenario 2

Klasifikasi	Learning rate	Error rate	epoch	MSE	Akurasi
Backpropagation Lavenberg-Marquadt	0,1	27,81%	7	0,1985	72,19%
	0,5	27,81%	8	0,1992	72,19%
	0,8	27,66%	16	0,1917	72,34%
<i>Backpropagation gradien descent</i>	0,1	27,96%	1000	0,2064	72,04%
	0,5	27,96%	1000	0,2023	72,04%
	0,8	29,03%	1000	0,2009	70,97%

5. Analisis

Tabel 4.1 Rata-Rata Akurasi

Klasifikasi	Learning Rate	Rata-rata MSE	Rata-rata Akurasi
<i>Backpropagation Lavenberg-Marquadt</i>	0.1	0,2020	72,01%
	0.5	0,2012	72,01%
	0.8	0,1979	72,21%
<i>Backpropagation gradien descent</i>	0.1	0,2012	71,81%
	0.5	0,2009	72,01%
	0.8	0,2003	72,01%

Berdasarkan tabel 4.5 rata-rata akurasi untuk setiap pengujian dapat dilihat bahwa pengujian di learning rate 0.8 memiliki tingkat akurasi tertinggi yaitu sebesar 72,21%. Hasil tersebut didapatkan dari rata-rata disetiap pengujian learning rate. Pada penggunaan learning rate yang besar pada metode backpropagation dapat mempengaruhi nilai MSE yang lebih rendah seperti pada pengujian di learning rate 0.8 dengan backpropagation lavenberg-marquadt dengan nilai MSE 0.1979. Pada backpropagation

lavenberg-marquadt dengan Pengujian 1 dan 2, menghasilkan performansi yang sangat baik. Dengan banyaknya data yang bisa di learning semakin akurat data akurasi yang dihasilkan. Seperti pada pengujian ke 2 dimana setiap hasil pengujian menghasilkan tingkat akurasi diatas 72%.

6. Kesimpulan

Berdasarkan hasil pengujian yang dilakukan menggunakan metode backpropagation menghasilkan performansi sistem yang baik tetapi mendapatkan error yang besar dengan nilai rata-rata kesalahan 28%. Hal ini disebabkan karena adanya data pengguna yang memiliki kemiripan karakteristik yang hampir sama pada kelas yang menyebarkan dengan yang tidak menyebarkan sehingga mempengaruhi pada pelatihan dan pengujian. Sehingga sistem tidak dapat mengklasifikasikan data secara keseluruhan dengan benar.

Metode backpropagation trainlm dan traingd dapat diimplementasikan dengan baik. Terlihat dari hasil pengujian yang menghasilkan nilai akurasi diatas 70% dengan 2 skenario yang dilakukan. Performansi terbaik didapat pada pelatihan dan pengujian di backpropagation trainlm dengan nilai rata-rata 72,21% Sedangkan pada backpropagation traingd memiliki akurasi yang baik namun membutuhkan waktu pelatihan yang relatif lama dengan mencapai batas maksimum epoch 1000.

Pada penggunaan learning rate yang besar pada metode backpropagation dapat mempengaruhi nilai MSE yang lebih rendah. Hal ini dapat dibuktikan pada 2 skenario pengujian yang telah dilakukan dan mendapatkan nilai MSE yang rendah dari learning rate dengan nilai 0,8 pada metode lavenberg-marquadt dengan nilai MSE 0.1979.

DAFTAR PUSTAKA

- [1] Situngkir, Hokky. (2011). "*Spread Of Hoax in Social Media*". Bandung : Department Computational Sociology.
- [2] Lind., P. G., da Silva, L. R., Andrade Jr., J. S., Herrmann, H. J. (2007). "*Spreading Hoax in Social Networks*". Physical Review E 76:33, 36117
- [3] S, Emilio., Iglesias, Carlos A., G, Mercedes., (2015). "A Novel Agent Based- Rumor Spreading Model in Twitter". Universidad Politecnica de Madrid.
- [4] Haykin, S. (1999) *Neural Networks: A Comprehensive Foundation*, Prentice Hall.
- [5] Suyanto., (2008). *Soft Computing*. Bandung :Teknik informatika
- [6] Herlina, Chaty., Bijaksana, Arif., Wibowo, Agung T. (2008). "*komunikasi Churn Pada telekomunikasi seluler dengan metode Jaringan Syaraf Tiruan*". Universitas Telkom.
- [7] Jaringan Syaraf Tiruan (Neural Networks). [online], available: <http://entin.lecturer.pens.ac.id> [Diakses 24 Oktober 2016].
- [8] Kristanto, A., (2004). "*Jaringan Syaraf Tiruan: Konsep Dasar, Algoritma dan Aplikasi*". Penerbit Gava Media, Yogyakarta.
- [9] L.J. Zhao, H.X. Cui, X.Y. Qiu, X.L. Wang, J.J. Wang, (2013) "*SIR rumor spreading model in Social Network*". Physica A 392 995–1003. Shanghai University.
- [10] Yann, LeCun., Bottou, Leon., Genevieve, B., K.B, Muller. "*efficient Backprop*". Research Department AT&T Labs-Research, 100 Schulz Driver
- [11] Twitter. (2013). <https://Support.Twitter.com/>. [Diakses 26 Oktober 2016].
- [12] Raymond W, J.Ryan S, Marco D,Richard B, David. B,J, T.S.Ugenne,"*A Network Application Programming Interface for Data Processing in Sensor Networks*", Department Rice University.
- [13] Shaw, A. K., Tsvetkova, M., & Daneshvar, R. (2011). "*The Effect of Hoax on Social Networks*". Journal Complexity.
- [14] Jimenez, Daniel A., Lin, Calvin, "*Dynamic Branch Prediction with Perceptrons*". Department of Computer Sciences. The University of Texas at Austin. Austin, TX 78712.
- [15] Arifien., Novie Fitriani, Arifin., Syamsul, Widjiantoro., Bambang Lelono. "*Prediksi kadar polutan menggunakan jaringan Syaraf tiruan (jst) untuk pemantauan Kualitas udara di kota surabaya*". Department of Engineering Physics, Faculty of Industrial Technology. ITS Surabaya Indonesia 60111.
- [16] Pant. Gautam., Srinivasan. Padmini., Menczer, Flippo. "*Crawling the Web*". Department of Management Sciences. Universit of Iowa, Iowa city, USA.