

Prediksi *Rating* Otomatis Berdasarkan *Review* Restoran pada Aplikasi *Zomato* dengan menggunakan *Extreme Learning Machine* (ELM)

Diajeng Tania Ananda Paramitha¹, Imam Cholissodin², Candra Dewi³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹taniaananda09@yahoo.com, ²imamcs@ub.ac.id, ³dewi_candra@ub.ac.id

Abstrak

Seiring berkembangnya zaman, perkembangan teknologi semakin berkembang. Salah satunya aplikasi untuk mencari informasi mengenai restoran di Jakarta yaitu *Zomato*. *Zomato* merupakan aplikasi yang menyediakan informasi dari berbagai restoran, fasilitas, harga, *review*, dan *rating* dari restoran tersebut. Masyarakat dapat memberikan *review* dan *rating* pada restoran tersebut. Data *review* tersebut berguna untuk pengguna sebelum ke restoran tersebut. Data *review* tersebut terkadang belum disertai *rating* sehingga membuat pemilik restoran mendapatkan masalah dalam mengklasifikasikan *review* kedalam *rating* untuk melakukan evaluasi kedepannya pada restoran tersebut. Pada penelitian ini membantu untuk mengklasifikasikan *review* kedalam *rating*. Pengujian metode ini menggunakan prediksi dengan metode *Extreme Learning Machine* (ELM). Proses prediksi ini menggunakan tahapan *pre-processing*, pembobotan kata dengan TF-IDF, dan perhitungan metode *Extreme Learning Machine* (ELM). Terdapat tahapan-tahapan pada metode ELM antara lain normalisasi, proses *training*, dan proses *testing*. Metode ELM ini menghasilkan akurasi sebesar 80,01% Dengan jumlah k yaitu 10 menggunakan *hidden neuron* sebanyak 25 dengan Interval bobot -0,5 hingga 0,5 fungsi aktivasi Sigmoid biner. Dapat disimpulkan bahwa metode ELM dapat menyelesaikan masalah prediksi dengan cukup baik.

Kata kunci: Prediksi, *rating*, *review*, *Extreme Learning Machine*.

Abstract

In this modern culture, technology advancement are growing better than we ever discovered before. One of the apps we use to search for information about restaurant in Jakarta are known as Zomato. Zomato is an application that provides various information about a restaurant from it facility, price, review, and rating. Users of The Zomato App can input various information that people haven't aware of about the restaurant into the app. Besides of inputting information into the app, Users of The Zomato App can also input a review and rating of a specific restaurant. The data review is used as an information about the restaurant for the potential customer from The Zomato App but sometimes the data review doesn't yet include a restaurant rating. This lack of misinformation will surely make the restaurant owner to occure some difficulties such as improving the restaurant services status for future outcomes. This research helps to classifying the review into the rating. Test protocol of this research are using a prediction with Extreme Learning Machine (ELM) Methods as it core. The prediction process however are build from a several steps such as pre-processing, word weighting with TF-IDF, and Extreme Learning Machine (ELM) Method calculations. Test result of The ELM parameter provides accuracy result 80,01% with k=10 amount hidden neuron 25 Interval weights -0.5 until 0,5 using function activation Sigmoid biner. We have come to conclusion were ELM method could positively solve the prediction problem exquisitely.

Keywords: *Rating*, *Prediction*, *Review*, *Extreme Learning Machine*

1. PENDAHULUAN

Zaman sekarang teknologi semakin canggih, semakin berkembangnya zaman. Pada era sekarang informasi mengenai berbagai hal

sangat mudah untuk didapatkan. Salah satunya untuk mencari restoran di kota-kota besar. Pencarian restoran biasanya dilakukan untuk memudahkan konsumen melihat menu yang ditawarkan, harga, dan fasilitas yang disediakan.

Sebelum memilih restoran biasanya konsumen dapat melihat *review* dari konsumen lain yang memberikan pengalamannya di sosial media.

Zomato merupakan salah satu pengembangan aplikasi *mobile* yang semakin canggih untuk membantu masyarakat mencari restoran berdasarkan kategori-kategori tertentu. Zomato dapat digunakan melalui *mobile* dan dapat juga diakses melalui *website* Zomato. Aplikasi tersebut memudahkan masyarakat khususnya di kota-kota besar untuk mencari restoran, *coffeeshop*, *live music*, dan lain sebagainya dari kota yang dipilih. Konsumen dapat memilih lokasi yang diinginkan, selain itu dapat melihat menu dari restoran tersebut, dapat memberikan *review*, dan masyarakat dapat melihat *rating* dari *review* pengalaman konsumen lain. Setiap *review* pada aplikasi Zomato juga telah diberikan *rating*. Menurut (Jong, 2011) *rating* merupakan gambaran dari kepuasan konsumen dalam memilih restoran. Dengan banyaknya data yang melimpah dari berbagai konsumen dan beberapa *review* juga belum adanya *rating* maka membuat pihak restoran mengalami kesulitan dalam mengklasifikasikan *review* kedalam *rating* dan mengetahui kekurangan dari restoran tersebut. Sehingga, penting adanya sistem untuk membantu mempercepat prediksi *rating* berdasarkan *review* konsumen agar restoran dapat melakukan evaluasi ke depannya.

Metode yang digunakan untuk memprediksi sesuatu telah banyak dilakukan dengan beragam implementasi yang mudah hingga yang rumit, Salah satunya yaitu *Extreme Learning Machine*. Penelitian sebelumnya yang dilakukan untuk memprediksi nilai tukar mata uang menggunakan ELM menghasilkan *Mean Square Error* sebesar 0,0000368 pada kurs jual dan 0,001596 pada kurs beli dengan 5 *input layer* dan 25 *hidden layer*.

Penelitian selanjutnya pada prediksi pemotongan *laser* menggunakan ELM dibandingkan dengan ANN dan *Genetic Algorithm* (GP). Hasil penelitian tersebut metode ELM memiliki nilai *error* terendah dibandingkan ANN dan GP.

Berdasarkan uraian dari penjelasan sebelumnya, membuktikan bahwa metode *Extreme Learning Machine* dapat mengatasi permasalahan dibandingkan metode lain untuk prediksi *rating* otomatis berdasarkan *review* restoran sebagai bahan evaluasi restoran kedepannya.

2. DASAR TEORI

2.1 Rating

Rating merupakan bagian terpenting yang akan dilihat oleh konsumen yaitu mencerminkan dengan benar baik atau buruknya dalam peringkat *rating*. Salah satunya pada penelitian ini menggunakan *review* beserta *rating* restoran. *Rating* dari *review* restoran dapat membantu konsumen dan pemilik restoran dalam mengetahui kualitas restoran, baik itu menu ataupun pelayanannya, layak atau tidak untuk dipilih. *Review* ini tidak hanya dibutuhkan oleh konsumen, tapi juga pemilik restoran. Pemilik restoran dapat melihat bagaimana respon dari konsumen. *Rating* yang digunakan yaitu *rating* 1 hingga *rating* 5, semakin rendah *rating* yang diberikan semakin kurang bagus begitupun sebaliknya jika *rating* yang diberikan tinggi maka semakin baik kualitasnya.

2.2 Pre-Processing

Pre-processing merupakan penghapusan kata dan selanjutnya untuk proses klasifikasi (Haddi, et al., 2013). Proses ini dilakukan seperti penghapusan kata-kata tidak memiliki arti dari dokumen tersebut. Menghilangkan kata tersebut membuat proses yang dilakukan semakin mudah untuk menjadikan informasi pada dokumen tersebut. Proses ini data disiapkan untuk dianalisis dalam *pre-processing* terdiri dari *case folding*, *filtering*, tokenisasi, *stopword removal* dan *stemming*. Hasil *pre processing* ini dilanjutkan untuk memberikan bobot setiap kata dan masuk ke tahap selanjutnya pada metode ELM.

2.2.1 Case Folding

Proses pertama dari *review* yaitu huruf kapital akan diubah menjadi huruf kecil (*lower case*) pada teks tersebut disebut *case folding*.

2.2.2 Filtering

Proses *Filtering* dengan menghilangkan selain huruf yaitu angka dan tanda baca yang terdapat dalam kalimat.

2.2.3 Tokenisasi

Proses pemecah dokumen menjadi beberapa kumpulan kata yang berasal dari proses *filtering*. Tokenisasi dilakukan untuk memisahkan per spasi.

2.2.4 Stopword Removal

Sstopword removal ialah menghilangkan kata-kata yang tidak penting seperti “dan”, “yang”, “di”, “ke”, “dari”. Dihilangkannya kata-kata *stopword* dikarenakan penggunaannya terlalu umum, dan lebih berfokus kepada kata-kata penting (Ganesan, 2015). *Stopword* pada penelitian ini berupa *stopword* pada data *mining*.

2.2.5 Stemming

Stemming yaitu dengan menghilangkan kata yang memiliki imbuhan yang terdapat dalam kata. Proses *stemming* merupakan pencarian kata dasar yang berbeda-beda dan memiliki beberapa jenis *stemming*. Penelitian ini menggunakan *stemming* sastrawi yang sederhana agar mudah dilakukan dokumentasi (Librian, 2017).

2.3 Pembobotan TF-IDF

Pembobotan kata TF-IDF adalah penilaian yang efektif dalam metode pembobotan setiap *term*. *Term Frequency-Inverse Document Frequency* untuk mengindikasikan seberapa unik terhadap kemunculan kata pada kumpulan dokumen (Nurjannah, 2016).

Term Frequency (TF) yaitu banyaknya frekuensi *term* pada dokumen. Nilai TF dapat menggunakan Persamaan 1.

$$tf_{t,d} = \begin{cases} 1 + \log_{10} tf_{t,d} & \text{if } tf_{t,d} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Setelah menghitung *term frequency* maka selanjutnya menghitung *idf_i* pada Persamaan 2.

$$idf_t = \log_{10} \frac{N}{df_t} \quad (2)$$

Selanjutnya menghitung TF-IDF menggunakan Persamaan 3.

$$W_{t,d} = tf_{t,d} \times idf_t \quad (3)$$

Keterangan:

- $tf_{t,d}$ = Frwkuensi munculnya kata t di dokumen d
- idf_t = Banyak dokumen yang mengndung term/token/kata
- N = Jumlah dkumen
- df_t = Dokumen yang memuat t
- $W_{t,d}$ = Bobot suatu *term*/token/kata

2.4 Jaringan Saraf Tiruan

JST dibuat untuk mengambil kelebihan dari otak. JST dikembangkan untuk membiatu tiruan seperti saraf manusia. (Jatmiko, et al.,2011)

JST merupakan informasi yang memiliki jenis serupa dengan neural biologis. Setiap JST memiliki karakteristik seperti, diantaranya:

- Pola hubungan antara elemen yakni *neuron*
- Penentuan nobot hubungan antar neuron
- Fungsi Aktivasi

2.5 Normalisasi

Normalisasi dilakukan ketika nilai input memiliki jarak yang bervariasi seperti puluhan hingga ribuan. Normalisasi sangat berguna untuk klasifikasi algoritme yang melibatkan jaringan saraf, atau jarak pengukuran seperti klasifikasi tetangga terdekat dan pengelompokan. Ada berbagai metode dalam melakukan normalisasi seperti *Z-Score*, *Decimal Scaling*, dan *Min-Max Normalizatio* (Jain & Bhandare, 2011). Berikut Persamaan 4 yaitu *Min-Max Normalization*.

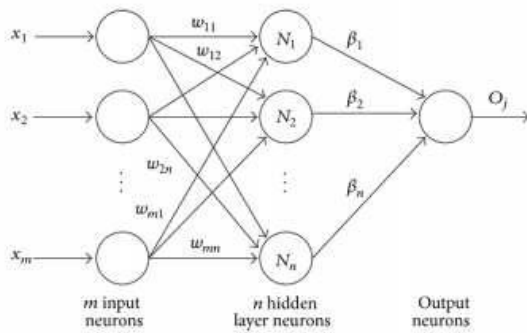
$$d' = \frac{d - \min(p)}{\max(p) - \min(p)} \quad (4)$$

Keterangan:

- d' = hasil normalisasi
- d = data yang di normalisais
- $\min(p)$ = nilai minimum dalam data
- $\max(p)$ = nilai maksimum dalam data

2.6 Extreme Learning Machine

Extrme Larning Machine memiliki sebutan yaitu *single hidden layer feedforward neural network* karena JST yang memiliki satu *hidden layer* (Huang, et al., 2006). Algoritme ini dilakukan agar menangani dalam hal *learning speed* dalam JST, karena selama ini JST memiliki *learning speed* rendah. Berikut arsitektur ELM pada Gambar 1.



Gambar 1. Asitektur ELM

2.6.1 Training Extreme Learning Machine

Tahapan-tahapan pada proses *training*, diantaranya:

1. Inisialisasi bias dan bobot dengan *random*
2. Inisialisasi *output hidden layer*

$$H_{init} = X_{training} \cdot W^T + b \quad (5)$$

Keterangan:

H_{init} = Hasil *output hidden layer*
 $X_{training}$ = Data *training*
 W^T = Transpose *input weight*
 b = Nilai bias

3. Aktivasi Sigmoid Biner pada Persamaan 11.
4. Inisialisasi *Moore Penrose*

$$H^+ = (H^T \cdot H)^{-1} \cdot H^T \quad (6)$$

Keterangan:

H^+ = matrik *Moore Penrose*
 H = matrik fungsi aktivasi
 H^T = matrik tranpose fungsi aktivasi
 $(H^T \cdot H)^{-1}$ = Matriks inverse

5. Inisialisasi *Output Weight*

$$\beta = H^+ \cdot Y \quad (7)$$

Keterangan:

β = nilai *output weight*
 H^+ = nilai *moore penrose*
 Y = nilai target

2.6.2 Testing Extreme Learning Machine

Tahapan-tahapan pada proses *testing*, diantaranya:

1. Inisialisasi bobot, bias, dan β dari proses *training* sebelumnya

2. Inisialisasi *Output Hidden Layer* pada Persamaan 8, sebagai berikut:

$$H_{init} = X_{testing} \cdot W^T + b \quad (8)$$

Keterangan:

H_{init} = Hasil *output hidden layer*
 $X_{testing}$ = Data *testing*
 W^T = Transpose *input weight*
 b = Bias

3. Fungsi aktivasi Sigmoid Biner pada Persamaan 11.
4. Prediksi (\hat{y})

$$\hat{y} = H \cdot \beta \quad (9)$$

Keterangan:

\hat{y} = Nilai *output layer*
 H = Nilai fungsi aktivasi
 β = Nilai *output weight*

5. Evaluasi

Akurasi =

$$\frac{\text{Data uji yang benar}}{\text{Semua data}} \times 100\% \quad (10)$$

2.7 Fungsi Aktivasi

Fungsi aktivasi yang diterapkan menurut (Srimuang, et al., 2015) yaitu:

1. Fungsi Aktivasi Sigmoid Binner

$$h(x) = \frac{1}{1 + \exp(-H_{init})} \quad (11)$$

2. Fungsi Aktivasi Sin

$$h(x) = \sin(H_{init}) \quad (12)$$

3. Fungsi Aktivasi Sigmoid Bipolar

$$h(x) = \frac{1 - \exp(-H_{init})}{1 + \exp(-H_{init})} \quad (13)$$

2.8 K-Fold Cross Validation

K-Fold Cros Validation ialah teknik validasi yang memahami kesesuaian data dengan sistem tersebut. Pada teknik ini data dibagi menjadi *k* bagian masing-masing *k* bagian memiliki jumlah yang sama. Misalkan menggunakan 5 *fold*, pertama menggunakan data *testing* pada *fold* pertama dan sisanya data latih. Selanjutnya berulang hingga *fold* terakhir, tujuan dari teknik ini untuk memberikan kesempatan dalam

melakukan uji terhadap data yang digunakan. (Kohavi, 2015).

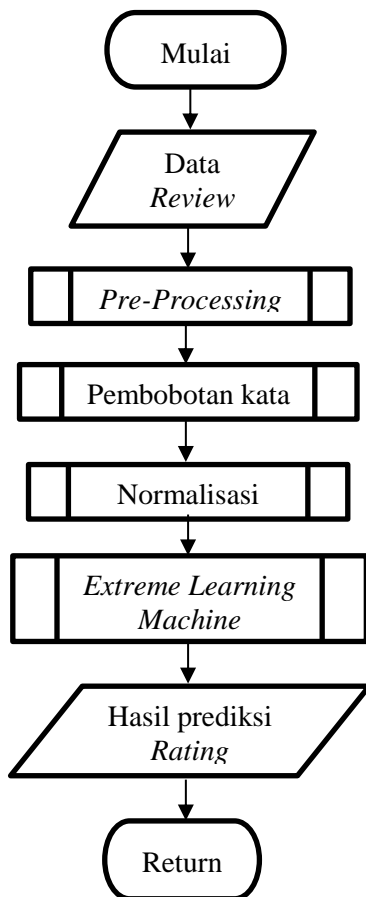
3. METODOLOGI PENELITIAN

3.1 Data Penelitian

Data yang digunakan yaitu *review* restoran beserta *rating* pada aplikasi Zomato sebanyak 150 data. Jumlah *rating* merupakan penjabaran dari masing-masing *rating* 1 hingga *rating* 5 dengan jumlah yang sama setiap *rating*.

3.2 Perancangan Algoritme

Proses alur pada prediksi *rating* berdasarkan *review* penelitian ini antara lain terdapat pada Gambar 2 ditunjukkan sebagai diagram alur dalam penelitian ini.



Gambar 2. Diagram alir

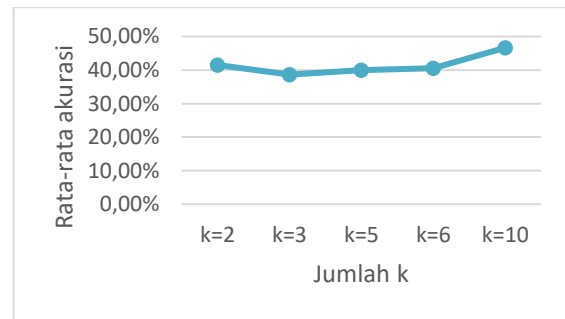
Langkah pertama yaitu data yang digunakan berupa data *review* dari aplikasi Zomato, selanjutnya data tersebut melakukan *pre-processing*. Setelah itu dilakukan TF-IDF, selanjutnya data tersebut di normalisasi, normalisasi yang digunakan yaitu *Min-Max*. Setelah proses normalisasi selesai maka

selanjutnya masuk ke dalam perhitungan metode ELM meliputi tahapan *training* dan tahapan *testing*. Setelah proses perhitungan testing selesai, didapatkan hasil prediksi *rating* tersebut.

4. HASIL DAN PEMBAHASAN

4.1 Pengujian K-Fold Cross Validation

Pengujian pertama bertujuan sebagai kesesuaian data yang digunakan dengan metode ELM. Pengujian ini menggunakan 150 data ketika $k=2$, $k=3$, $k=5$, $k=6$, dan $k=10$ dengan fungsi aktivasi sigmoid biner dan jumlah *hidden* sebesar 50, interval bobot yang digunakan yaitu -0,5 hingga 0,5.

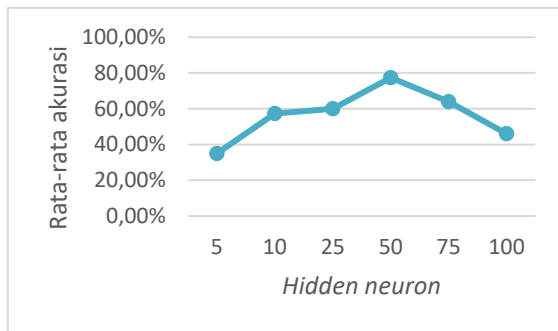


Gambar 3. Pengujian K-Fold

Pada Gambar 3 menunjukkan grafik pengujian *k fold cross validation*, dilihat ketika $k=2$, $k=3$, $k=5$, $k=6$, dan $k=10$ hasil tersebut tidak berbeda jauh walaupun menggunakan nilai bobot dan bias secara *random*. Hasil rata-rata tertinggi dari 5 percobaan pada $k=10$ yaitu sebesar 46.66% dan hasil rata-rata dari semua nilai k yaitu 41.48%. Selain itu hasil rata-rata terendah yaitu terdapat pada $k=3$ yaitu 38.65%. Hal tersebut menunjukkan bahwa metode ELM dengan nilai k semakin tinggi maka semakin banyak pula jumlah data *training* yang digunakan sehingga *term* yang dihasilkan semakin bervariasi dan hasil akurasi semakin baik.

4.2 Pengujian Jumlah Hidden Neuron

Pengujian *hidden neuron* untuk memahami jumlah *hidden neuron* pada evaluasi yang dihasilkan. Pengujian ini dilakukan menggunakan parameter terbaik pada pengujian *k-fold* dengan 5 kali percobaan. *Hidden neuron* yang dilakukan sebanyak 5, 10, 25, 50, 75, dan 10. Gambar 4 menunjukkan hasil dari pengujian tersebut.

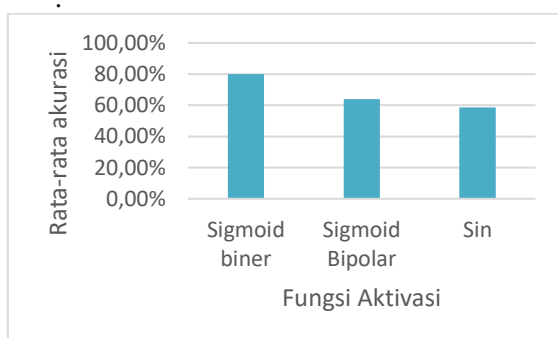


Gambar 4. Pengujian jumlah *hidden neuron*

Gambar 4 menghasilkan akurasi yang bervariasi, dapat dilihat ketika jumlah hidden neuron sebesar 75 maka mengalami penurunan hingga jumlah hidden neuron sebesar 100 semakin menurun. Hasil akurasi tertinggi ketika jumlah *hidden neuron* sebanyak 50 yaitu sebesar 77,34%. Rata-rata akurasi terendah dari 5 percobaan terdapat pada jumlah *hidden layer* sebanyak 5 yaitu sebesar 34,99. Sehingga dapat disimpulkan bahwa semakin besar jumlah *hidden neuron* tidak selalu menghasilkan akurasi yang baik.

4.3 Pengujian Fungsi Aktivasi

Pengujian ketiga yaitu pengujian aktivasi untuk mencari fungsi aktivasi mana yang tepat pada permasalahan ini. Pengujian ini menggunakan hasil optimal pada pengujian sebelumnya. Fungsi aktivasi yang dilakukan berupa Sigmoid Biner, Bipolar, dan Sin. Pada Gambar 5 menunjukkan grafik hasil pengujian tersebut.



Gambar 5. Pengujian Aktivasi

Gambar 5 dapat dilihat bahwa setiap percobaan menghasilkan rata-rata akurasi yang bervariasi. Maka akan terlihat dalam pengujian fungsi aktivasi yang terendah pada fungsi aktivasi Sigmoid Bipolar dan Sin. Namun untuk akurasi tertinggi ditunjukkan pada fungsi aktivasi Sigmoid Biner sebesar 80,01% dalam 5 kali percobaan. Selain itu dapat diketahui rata-

rata hasil akurasi terendah sebesar 58,67% pada fungsi aktivasi Sin. Dapat disimpulkan bahwa masing-masing fungsi aktivasi memiliki karakteristik yang berbeda-beda, seperti hasil penelitian ini fungsi sigmoid biner sesuai dengan data yang digunakan yaitu data biner.

5. KESIMPULAN

Berdarkan hasil dari pengujian serta analisis dari pengujian tersebut maka didapatkan kesimpulan, diantaranya:

1. Metode *Extreme Learning Machine* (ELM) mampu menyelesaikan permasalahan prediksi *rating* otomatis berdasarkan *review* restoran pada aplikasi Zomato dengan akurasi cukup baik. Proses implementasi yang dilakukan yaitu *pre-processing* yang terdiri dari *case folding*, *filtering*, tokenisasi, *stopword removal*, dan *stemming*. Tahap selanjutnya adalah ekstraksi fitur dengan menggunakan pembobotan TF-IDF, lalu normalisasi, dan selanjutnya menggunakan *Extreme Learning Machine*.
2. Berdasarkan dari hasil pengujian prediksi *rating* otomatis berdasarkan *review* restoran pada aplikasi Zomato menggunakan *Extreme Learning Machine* dengan 150 data yang terdiri dari 135 data *training* dan 15 data *testing* saat $k=10$. Sehingga menggunakan *hidden layer* sebanyak 50 dengan fungsi aktivasi Sigmoid Biner. Pada metode *Extreme Learning Machine* (ELM) menghasilkan akurasi sebesar 80,01%. Nilai akurasi tersebut dinilai cukup baik untuk menyelesaikan permasalahan ini.

6. DAFTAR PUSTAKA

- Ganesan, K., 2015. A Brief Note on Stop Words for Text Mining and Retrieval. Available at: <http://www.text-analytics101.com/2014/10/allabout-stop-words-for-text-mining.html>.
- Haddi, E., Liu X., dan Shi, Y. (2013). The Role of Text Pre-Processing in Sentiment Analysis. *Procedia Computer Science*, pp.26-23.
- Huang, G. B., Zhu, Q. Y. & Siew, C. K., 2006. *Extreme Learning Machine : Theory and Application*. Elsevier (*Scient Direct*), Volume 70, pp. 489-501.
- Jain, Y.K. & Bhandare, S. K., 2011. Min Max Normalization Based Data Perturbation

- Method For Privacy Protection. *Computer Science & Engineering*, 2(VIII), pp. 45-50
- Jatmiko, W., Mursanto, P., Fajar, M., Tawakal, M.I., Trianggoro, W., Rambe, R.S., Ramadhan, F. and Arief, 2011. *Implementasi Berbagai Algoritme Neural Network Dan Wavelet Pada Field Programmable Gate Array*.
- Jong, J., 2011. Predicting Rating with Sentiment Analysis. [Online] tersedia di: <https://scholar.google.co.id/> [diakses 2 September 2018]
- Kohavi, R., 2015. A Study of Cross Validation and Bootstrap for Accuracy Estimation and Model Selection. *Computer Science*.
- Librian, A., 2017. High quality stemmer library for Indonesian Language (Bahasa). [Online] tersedia di: <https://github.com/sastrawi/sastrawi> [diakses 12 November 2018].
- Manning, C. D., Raghavan, P. & Schütze, H., 2009. *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Nurjannah, M., Hamdani, H. and Astuti, I.F., 2016. Penerapan Algoritme Term Frequency-Inverse Document Frequency (TF-IDF) untuk Text Mining. *Jurnal Informatika Mulawarman (JIM)*, 8(3), pp.110–113.
- Srimuang, W, Intarashonchun & S, 2015. Classification Model of Network Intrusion Using Weighted Extreme Learning Machine. *International Joint Conference on Computer Science And Software Engineering (JCSSE)*, Volume 12.