

# Klasifikasi Sentimen *Review* Produk Otomotif Menggunakan *Back Propagation Neural Network*

Fuad Ash Shiddiq  
Fakultas Informatika  
Universitas Telkom  
Bandung, Indonesia  
fuadashshiddiq@students.telkomuniversity.ac.id

Said Al Faraby  
Fakultas Informatika  
Universitas Telkom  
Bandung, Indonesia  
saidalfaraby@telkomuniversity.ac.id

Adiwijaya  
Fakultas Informatika  
Universitas Telkom  
Bandung, Indonesia  
adiwijaya@telkomuniversity.ac.id

Abstrak- Perkembangan teknologi sekarang menuntut untuk dapat mengikuti perubahan jaman. Salah satunya dalam transaksi jual-beli yang dilakukan secara online. Sebuah toko online dapat menjual berbagai jenis barang misalnya barang elektronik, mainan, bahkan barang otomotif. Dengan adanya transaksi secara digital banyak pembeli yang ingin mengetahui bagaimana ulasan atau *review* tentang barang yang dijual di toko online dari pembeli yang sebelumnya telah membeli barang tersebut. Banyaknya *review* yang diberikan oleh konsumen menimbulkan data *review* untuk suatu produk di toko online menjadi sangat besar. Untuk melakukan klasifikasi data yang besar tersebut diperlukan sistem yang terotomatisasi. Dalam penelitian ini, sistem yang dibangun menggunakan metode *Back-propagation Neural Network* untuk melakukan klasifikasi data *review*. Hasil yang didapatkan dengan akurasi paling tinggi adalah dengan jumlah *hidden layer* : 1000, *epoch*:400, dan *learning rate* : 0.2 dengan akurasi sebesar 60%. Untuk pengujian tanpa menggunakan proses *stopword removal* didapatkan dengan akurasi paling tinggi dengan jumlah *hidden layer* : 1000, *epoch*:400, dan *learning rate* : 0.2 dengan akurasi sebesar 56.7%. Hal ini dikarenakan proses *stopword removal* dapat mengurangi *noise* dalam sebuah data dan mempercepat proses klasifikasi.

**Kata kunci** : *review*, klasifikasi, *back-propagation*, *stopword removal*

## I. PENDAHULUAN

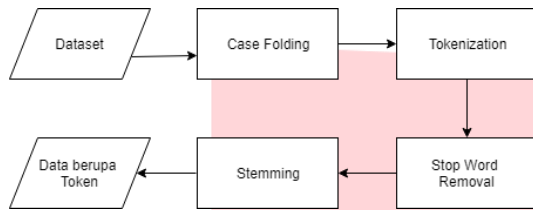
Saat ini sebelum membeli suatu barang di sebuah toko online kita biasanya mempertimbangkan pendapat/opini orang lain yang sudah terlebih dahulu membeli barang elektronik tersebut, begitu juga sebelum menonton film di bioskop atau bahkan saat akan memilih pada pemilu. Jaman dulu kita mendapatkan informasi dari sahabat, teman, saudara

atau dari laporan konsumen. Tetapi jaman telah berubah sekarang adalah jaman internet, orang-orang dengan mudahnya terhubung satu dengan yang lain. Banyak sekali cara untuk orang-orang menyampaikan pendapatnya melalui internet seperti lewat forum, blog atau kolom *review* produk pada toko online. *Review* ini membantu kedua belah pihak yaitu konsumen dan produsen. Untuk konsumen berguna untuk mendapatkan gambaran dari sebuah produk atau layanan dan untuk produsen itu sendiri adalah untuk meningkatkan kualitas dari produk atau layanan mereka. Banyak orang-orang yang berkeinginan untuk membagikan pendapat mereka atau hanya membagikan pengalaman mereka tentang sebuah produk atau layanan lewat internet, maka akan menghasilkan data yang sangat besar. Data yang sangat besar ini dibutuhkan sebuah sistem otomatis untuk mengolahnya. Sistem otomatis (Sentiment Classification System) ditujukan untuk melakukan proses analisis, meringkas, dan mengklasifikasi data. Terdapat banyak metode pendekatan untuk melakukan klasifikasi terhadap data *review* tersebut, yaitu Support Vector Machine(SVM), Maximum Entropy Modelling(MEM), Naive Bayes, Jaringan Saraf Tiruan(ANN) dan lain lain [1].

Dalam penelitian Tugas Akhir ini, untuk mengklasifikasi *review* dari produk otomotif pada toko online Amazon digunakanlah metode *Back Propagation Neural Network* untuk melakukan klasifikasi sentimen dalam bentuk rating pada tingkat dokumen. Untuk setiap input yang diterima, setiap neuron digunakan untuk mengaktifkan jaringan bersamaan dengan bobot untuk melatih sistem sesuai dengan aturan untuk melakukan prediksi. *Feed-Forward* dan *Back-Propagation* pada ANN membantu untuk melatih jaringan syaraf dan mempelajari statistik [2]. Jamet et. al [3] mengusulkan untuk menggunakan fungsi aktivasi yang dapat digunakan pada ANN seperti fungsi sigmoid yang dapat meningkatkan kemampuan sebuah jaringan syaraf.

## II. KLASIFIKASI TEKS

Klasifikasi teks merupakan cara untuk mengkategorikan teks ke dalam kelas yang sudah ditentukan sebelumnya [4]. Masalah utama pada klasifikasi ini bergantung pada kata-kata yang muncul di dokumen yang mempengaruhi fitur dan dimensi sehingga frekuensi kata berpengaruh terhadap hasil dari klasifikasi. Sebelum proses klasifikasi ini diperlukan merubah data teks menjadi token.



Gambar 1 Preprocessing

Berikut ini penjelasan dari gambar diatas :

- *Case Folding*, proses pembersihan data dimana seluruh kalimat diubah menjadi huruf kecil [5]

Data asli	The build quality is good. The included cables are of...
Setelah case folding	the build quality is good. the included cables are of...

- *Tokenization*, proses yang dilakukan sistem untuk menjadikan sebuah kalimat menjadi lebih bermakna dengan cara memecah kalimat tersebut menjadi kata-kata atau frase-frase [5]

Setelah case folding	the build quality is good. the included cables are of...
Hasil tokenisasi	the-build-quality-is-good-the-included-cables-are-of...

- *Stop Word Removal*, proses menghilangkan kata-kata yang

dianggap tidak penting seperti “the, of, a”. jika termasuk dalam daftar stop list maka kata-kata tersebut akan dihapus.[6]

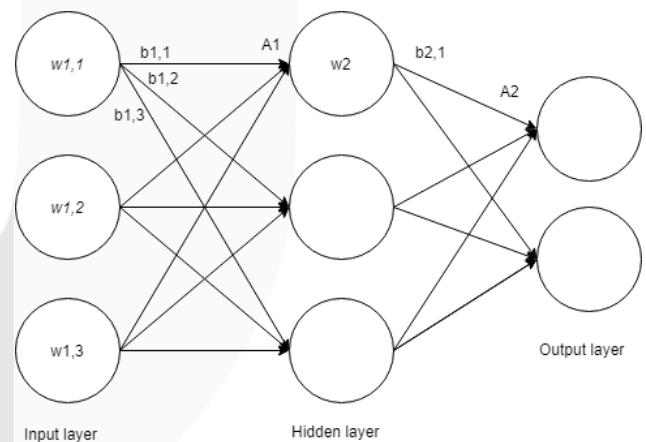
Hasil Tokenisasi	the-build-quality-is-good-the-included-cables-are-of...
Hasil <i>stopword removal</i>	build-quality-good-included-cables-

- *Stemming*, proses mengubah kata-kata yang berimbuhan menjadi bentuk aslinya[7].

Hasil <i>stopword removal</i>	build-quality-good-included-cables-
Hasil <i>stemming</i>	include-cable-

## III. ARTIFICIAL NEURAL NETWORK

*Neural Network* adalah salah satu pendekatan komputasi yang didasarkan pada koleksi besar dari *neural unit* yang memodelkan bagaimana cara otak memecahkan masalah dengan kluster yang besar dari biological neuron yang tersambung dengan akson. *Artificial Neural Network* adalah interkoneksi grup dari node-node.



Gambar 1 Artificial Neural Network

Untuk menghitung nilai setiap A1 dimana A1 adalah nilai sebuah node, yaitu dengan cara menentukan

besar  $w$  diawal ditentukan dengan cara melakukan random nilai bobotnya, begitu juga dengan nilai  $b$  yang didapatkan dari random nilainya. Untuk menghitung nilai  $A_1$  dengan menggunakan rumus sigmoid sebagai berikut :

$$A_1 = \frac{1}{1 - \exp^{1-v}}$$

Dimana nilai  $v$  didapatkan dari perhitungan :

$$v = w_1 \times \text{input} + b_1$$

Untuk mencari nilai dari  $A_2$  dengan perhitungan sebagai berikut :

$$A_2 = \frac{1}{1 - \exp^{1-v_2}}$$

Dengan nilai  $v_2$  yang didapatkan dari :

$$v_2 = w_2 \times A_1 + b_2$$

Dan seterusnya untuk layer *hidden* selanjutnya hingga didapatkan nilai dari outputnya

#### IV. TF-IDF FEATURE WEIGHTING

TF-IDF merupakan proses ekstraksi fitur yang digunakan untuk mengubah atribut *string* menjadi bobot yang memiliki houngan penting dengan *corpus*. Nilai dari TF-IDF ditinjau dari seringnya kata yang muncul dalam sebuah dokumen namun diimbangi dengan pengulangan kata pada *corpus*[8]. TF atau *Term Frequency* adalah ukuran seberapa sering kemunculan sebuah *term* dalam sebuah dokumen.. Perhitungannya sebagai berikut

$$tf = f_{t,d} / \sum_{t \in d} f_{t',d}$$

$tf$  = term frekuensi,  $f_{t,d}$  = frekuensi sebuah term dalam satu dokumen,  $\sum_{t \in d} f_{t',d}$  = total term pada sebuah dokumen

Sedangkan IDF atau *Inverse document frequency* berfungsi untuk mengurangi bobot suatu term jika kemunculan banyak tersebar di seluruh koleksi dokumen. Dokumen frekuensi adalah seberapa banyak term muncul diseluruh dokumen yang diperiksa. IDF =  $\log_{10}$  (Banyaknya dokumen/jumlah dokumen yang mengandung *term* tersebut) dengan rumus idf sebagai berikut.

$$idf_t = \log_{10} \left( \frac{N}{df_t} \right)$$

idf = inverse document frequency, N = banyak dokumen, df=jumlah dokumen yang mengandung term

Bobot dari sebuah *term* dihitung menggunakan ukuran tf-idf dengan persamaan

$$w_{t,d} = (1 + \log tf_{t,d}) \times \log_{10} \left( \frac{N}{df_t} \right)$$

## V. SISTEM

### A. Data

Dataset yang digunakan dalam penelitian ini adalah data *review* produk otomotif berbahasa inggris yang didapatkan dari toko online Amazon.com. Data yang digunakan berjumlah 450 dokumen dengan pembagian rating(1-5) secara acak. Perbandingan data latih dan uji yang digunakan sebesar 400:50, dengan jumlah *term* total sebanyak 5746 *term*.

### B. Sistem

Sistem akan dijelaskan sebagai berikut:

- Dataset akan dilakukan *preprocessing*, ada 4 langkah *preprocessing* yaitu *casefolding*, *tokenization*, *stopword removal*, *stemming*.
- Setelah proses *preprocessing* menghasilkan data yang berupa token-token dilakukan proses pembobotan fitur menggunakan TF-IDF.
- Data yang sudah diberikan bobot selajutnya akan dipisahkan antara data latih dan data uji.
- Setelah dipisahkan data-datanya dilakukan proses klasifikasi menggunakan *backpropagation neural network* untuk data latih maupun data uji.
- Setelah proses klasifikasi didapatkan lah akurasi dari klasifikasi tersebut yaitu dengan membandingkan prediksi rating dengan rating yang sudah ada pada datasetnya.

## VI. EVALUASI

### A. Hasil Pengujian

Untuk percobaan pada tabel 3 dilakukan dengan jumlah *hidden layer* 1000, *epoch* 100, *learning rate* 0.2. Dilakukan percobaan sebanyak 3 kali dan didapatkan akurasi sebesar 57.3 %..

Tabel 1 percobaan 1 menggunakan *stopword removal*

Percobaan ke-	<i>Hidden</i>	<i>epoch</i>	<i>Learning rate</i>	akurasi
1	1000	100	0.2	58 %
2	1000	100	0.2	56 %
3	1000	100	0.2	58 %
Rata-rata				57.3 %

Tabel 2 Percobaan 2 menggunakan *stopword removal*

Percobaan ke-	<i>Hidden</i>	<i>epoch</i>	<i>Learning rate</i>	akurasi
1	1000	400	0.2	60%
2	1000	400	0.2	60%
3	1000	400	0.2	60%
Rata-rata				60 %

Berdasarkan hasil dari percobaan diatas didapatkan akurasi paling besar menggunakan *parameter* jumlah *hidden layer* 1000, dengan 400 *epoch* dan 0.2 *learning rate* yaitu sebesar 60%.

Untuk percobaan tanpa *stopword removal* pada tabel 5 dilakukan dengan jumlah *hidden layer* 1000, *epoch* 100, *learning rate* 0.2. Dilakukan percobaan sebanyak 3 kali dan didapatkan akurasi sebesar 56%.

Tabel 3 Percobaan 1 tanpa *stopword removal*

Percobaan ke-	<i>Hidden</i>	<i>epoch</i>	<i>Learning rate</i>	akurasi
1	1000	100	0.2	56 %
2	1000	100	0.2	56 %
3	1000	100	0.2	56 %
Rata-rata				56 %

1	1000	100	0.2	56 %
2	1000	100	0.2	56 %
3	1000	100	0.2	56 %
Rata-rata				56 %

Tabel 4 Percobaan 2 tanpa *stopword removal*

Percobaan ke-	<i>Hidden</i>	<i>epoch</i>	<i>Learning rate</i>	akurasi
1	1000	400	0.2	56%
2	1000	400	0.2	56%
3	1000	400	0.2	58%
Rata-rata				56.7 %

### B. Analisis Hasil

Didapatkan hasil dari percobaan tanpa menggunakan *stopword removal*, dapat dilihat bahwa hasil yang didapatkan dibandingkan dengan yang menggunakan *stopword removal* lebih besar akurasi dari yang menggunakan *stopword removal* yaitu sebesar 57.3% untuk *epoch* 100 dan 60 % untuk *epoch* 400 dibandingkan yang tidak menggunakan *stopword removal* yaitu sebesar 56% untuk *epoch* 100 dan 56.7% untuk *epoch* 400.

## VII. KESIMPULAN

Berdasarkan hasil pengujian dan analisis yang telah dilakukan sebelumnya, maka dapat diambil kesimpulan yaitu, sistem yang dibangun menggunakan *backpropagation neural network* dapat menghasilkan akurasi sebesar 57.3% untuk *epoch* sebesar 100, 1000 *hidden layer* dan *learning rate* sebesar 0.2 sedangkan untuk *epoch* sebesar 400, 1000 *hidden layer* dan *learning rate* sebesar 0.2 didapatkan akurasi sebesar 60 %. Penggunaan *stopword removal* mempengaruhi hasil akurasi yakni dengan perbedaan sebesar 1.3 % dan 3.3%.

## DAFTAR PUSTAKA

- [1] Vimalkumar B. Vaghela, Bhumika M. Jadav  
*Analysis of Various Sentiment Classification Techniques, International Journal of Computer Applications (0975-8887), Volume 140-No 3, April 2016.*
- [2] A. Mudinas, D. Zhang and M. Levene,  
*Combining lexicon and learning based approaches for concept-level sentiment analysis*, in Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, 2012.
- [3] Hinton, *Reducing the dimensionality of data with neural networks*, Science, vol. 313, pp. 504-507, 2006..
- [4] C. C. Aggarwal. A. *SURVEY OF TEXT CLASSIFICATION ALGORITHM*. Springer, 2012
- [5] G. Grefenstette. *Syntactic Worldclass Tagging*. Springer Netherlands, 1999.
- [6] J. L., A. Rajaraman. *Mining of Massive Datasets*., 2014
- [7] H.S. Christopher D. Manning, Prabhakar Raghavan, *an Introduction to Information Retrieval*. Cambridge, 2009
- [8] J. Han. *Data Mining : Concept dan Technique Second Edition*. 2006
- [9] Pranali Borele, Dilipkumar A. Borikar, *An Approach to Sentiment Analysis Using Artificial Neural Network with Comparative Analysis of Different Techniques*, IOSR Journal of Computer Engineering (IOSR-JCE), 2016
- [10] Ms. Kranti Vithal Ghag, Dr. Ketan Shah, *Comparative Analysis of Effect Stopwords Removal on Sentiment Classification*, IEEE International Conference on Computer, Communication and Control, 2015