

IMPLEMENTATION OF BACKPROPAGATION NEURAL NETWORK IN SENTIMENT ANALYSIS ON TWITTER TO PUBLIC FIGURES

Achmad Safruddin¹, Arief Hermawan², Adityo Permana Wibowo^{3,*}

^{1,2}Program Studi Magister Teknologi Informasi

³Program Studi Informatika

^{1,2,3}Universitas Teknologi Yogyakarta

Jl. Siliwangi (Ringroad Utara), Jombor, Sleman, Yogyakarta

Email Korespondensi : *adityopw@uty.ac.id

Received: November 18, 2020; Accepted : November 26, 2020; Published : November 27, 2020

Abstract

Sentiment analysis is a process for identifying or analyzing people's opinions on a topic. Sentiment analysis analyzes each word in a sentence to find out the opinions or sentiments expressed in the sentence. The opinions expressed can be in the form of positive or negative opinions. Twitter is one of the most popular social media in Indonesia. Twitter users always discuss various kinds of topics every day. One of the things discussed on Twitter and which has become a trending topic several times is about public figures. This study discusses the analysis of positive or negative sentiments towards public figures based on tweet data carried out by text processing. The results of text processing are classified using a backpropagation neural network. Tests were carried out using 69 test data, resulting in an accuracy of 62.3%, with 43 correct classification results.

Keywords: *Sentiment Analysis, Public Figures, Backpropagation, Classification*

1. Latar Belakang Masalah

Tahun 2020, tren pengguna internet di Indonesia sebanyak 175,4 juta sedangkan pengguna aktif media sosial di Indonesia sebanyak 160 juta [1]. Beberapa platform media sosial yang aktif digunakan di Indonesia salah satunya adalah Twitter. Bulan Januari 2020 Twitter menempati urutan ke 4 atau sebanyak 58% media sosial yang sering diakses oleh pengguna di Indonesia [1]. Banyaknya pengguna twitter dapat dimanfaatkan untuk mengetahui sentimen masyarakat terhadap tokoh publik yang ada di Indonesia. Twitter salah satu platform media sosial yang menyediakan fitur untuk berbagi pesan menggunakan teks pendek sebanyak 280 karakter atau yang biasa disebut *tweet* [2].

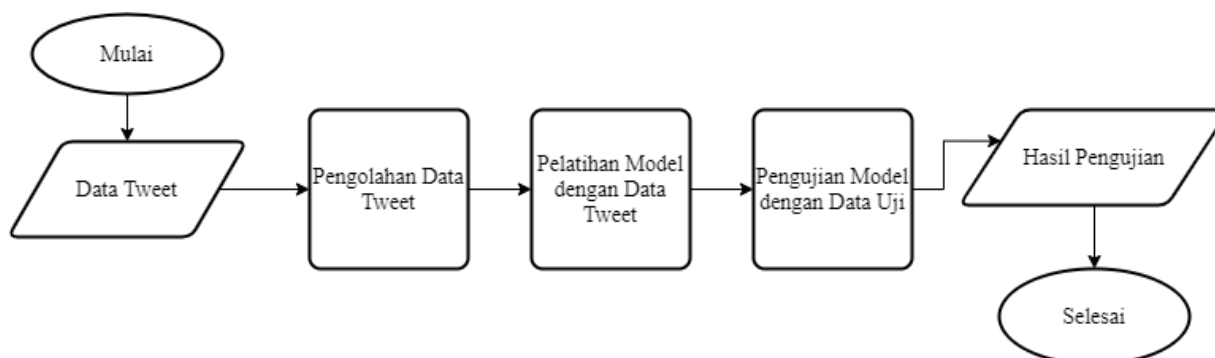
Tweet yang ditulis oleh pengguna Twitter memiliki sentimen tersendiri terhadap sebuah obyek seperti tokoh publik, peristiwa, benda, dan lain sebagainya. Sentimen tersebut biasanya terdiri dari sentimen positif, netral, dan negatif yang ditentukan secara manual oleh pembacanya sendiri. Penentuan tersebut biasanya disebut juga dengan *labeling*. Seiring berkembangnya teknologi, untuk melakukan analisis sentimen bisa menggunakan salah satu fungsi yaitu *Text Mining* [2][3]. Kumpulan *tweet* yang didapatkan dari twitter menggambarkan pola data yang tidak terstruktur sementara untuk melakukan proses *text mining* diperlukan data yang terstruktur, sehingga dengan demikian didalam *text mining* akan ada proses *crawling*, *text preprocessing*, *stemming*, dan klasifikasi [4].

Penelitian ini membahas tentang klasifikasi tokoh publik di Indonesia berdasarkan *tweet* yang didapatkan dari Twitter. Klasifikasi menggunakan Jaringan Syaraf Tiruan Backpropagation atau *Backpropagation Neural Network* (BNN). BNN merupakan versi modifikasi dari jaringan syaraf tiruan yang menambahkan propagasi error dengan bergerak mundur (dari lapisan keluaran menuju lapisan masukan). Penggunaan propagasi balik dapat mempercepat konvergensi sehingga model dengan jumlah data input yang banyak dapat dengan cepat didapatkan [5]. Kelebihan pada metode BNN tersebut menjadikan metode ini dapat digunakan untuk pengklasifikasian sentimen pada *tweet* pengguna Twitter terhadap sebuah obyek seperti tokoh publik.

Penelitian yang membahas tentang analisis sentimen sudah banyak dilakukan antara lain [5][6], sedangkan penggunaan Jaringan Syaraf Tiruan *Backpropagation* juga sudah banyak dilakukan mulai dari penggunaan untuk klasifikasi [7] [8] sampai dengan prediksi [9] [10].

2. Metodologi Penelitian

Penelitian dilakukan mulai dari *crawling* data Twitter, kemudian pengolahan data Twitter yang terdiri dari pelabelan, pembersihan *text* dari *emoticon* dan *hashtag*, data yang sudah bersih selanjutnya berturut-turut dilakukan proses *case folding*, *tokenizing*, dan terakhir *stemming*. Data yang sudah bersih, selanjutnya digunakan untuk pelatihan di BNN untuk membuat model analisis sentimen. Model yang sudah jadi kemudian akan diuji untuk menghasilkan model yang terbaik. Alur tahapan penelitian seperti terlihat pada Gambar 1.



Gambar 1. Alur Tahapan Penelitian

2.1 Data Penelitian

Penelitian ini menggunakan 158 data untuk melatih jaringan syaraf tiruan agar dapat mengenali sentimen positif atau negatif pada suatu *tweet*. Data yang digunakan merupakan data *tweet text* dengan bahasa Indonesia yang mengarah atau berhubungan kepada tokoh publik. *Tweet* dalam bentuk gambar, *video*, dan atau suara tidak digunakan dalam penelitian ini. *Crawling* data *tweet* menggunakan *Application Programming Interface* (API) yang disediakan oleh Twitter berdasarkan tagar (#) terhadap tokoh publik yang dikehendaki atau nama akun tokoh. Data yang dikirimkan oleh server memiliki format json (*javascript object notation*), kemudian harus diubah menjadi bentuk *dataframe*.

Beberapa contoh pernyataan positif dan negatif dari masyarakat terhadap tokoh Rocky Gerung terdapat pada Gambar 2. Sedangkan pernyataan positif masyarakat terdapat pada Gambar 3.



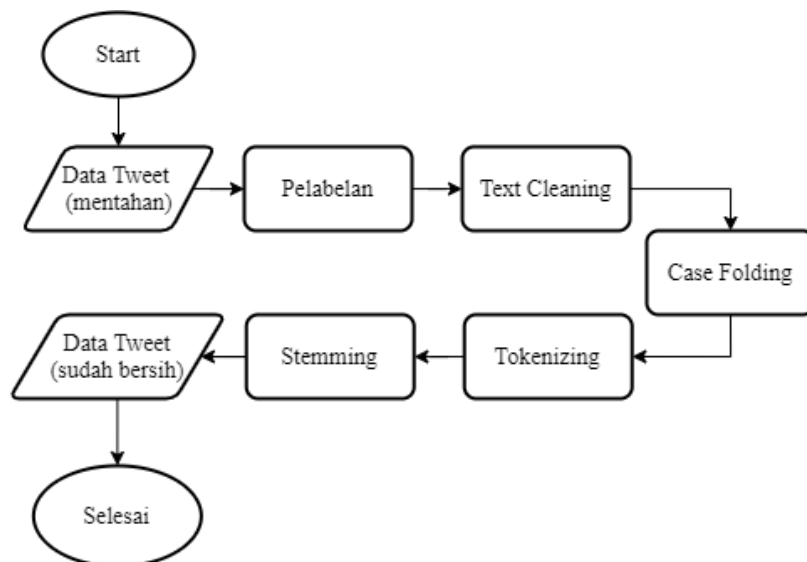
Gambar 2. Pernyataan Negatif pada Twitter



Gambar 3. Pernyataan Positif pada Twitter

2.2 Pengolahan Data Tweet

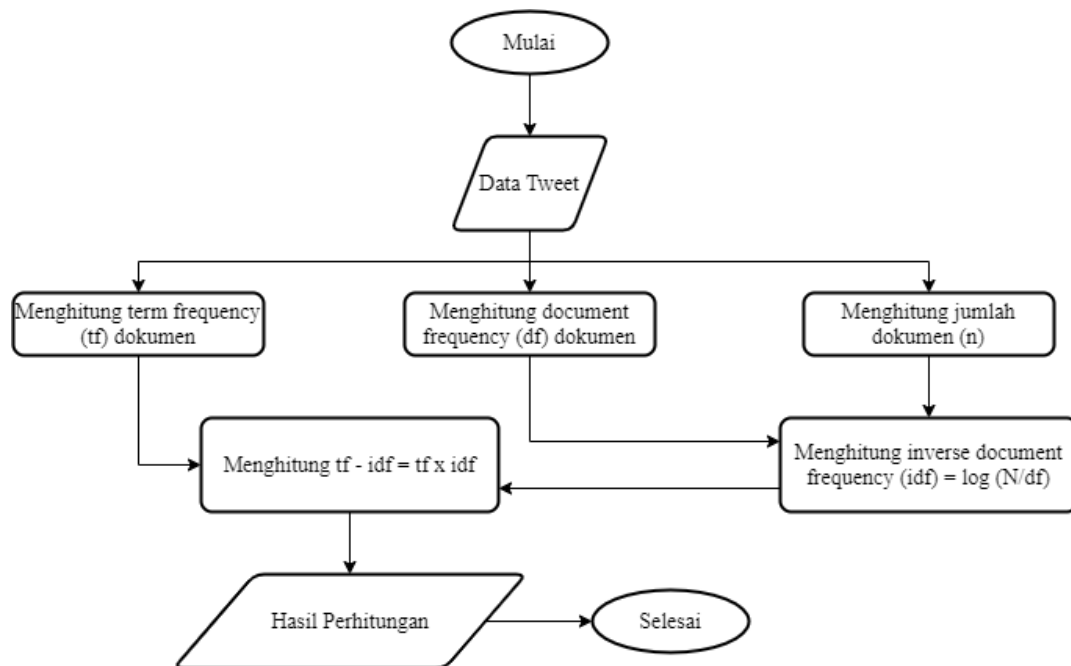
Data yang diambil dari Twitter akan diproses terlebih dahulu untuk dibersihkan dari *emoticon*, *hashtag*, dan *mention* sehingga benar-benar hanya berupa kata/kalimat. Beberapa proses yang dilakukan mulai dari pelabelan, pembersihan kata, *Case Folding*, *Tokenizing*, dan *Stemming* [4]. Alur pengolahan data seperti terlihat pada Gambar 4.



Gambar 4. Alur Pengolahan Data Twitter

2.3 Pembobotan *Term Frequency – Inverse Document Frequency (TF-IDF)*.

Setelah data *tweet* sudah dibersihkan, selanjutnya dilakukan perhitungan bobot setiap kata menggunakan perhitungan *Term Frequency – Inverse Document Frequency (TF – IDF)*. TF – IDF merupakan algoritma yang digunakan untuk menentukan bobot dari suatu kata (t) pada suatu dokumen (d) [11]. Alur pembobotan TF-IDF seperti terlihat pada Gambar 5.

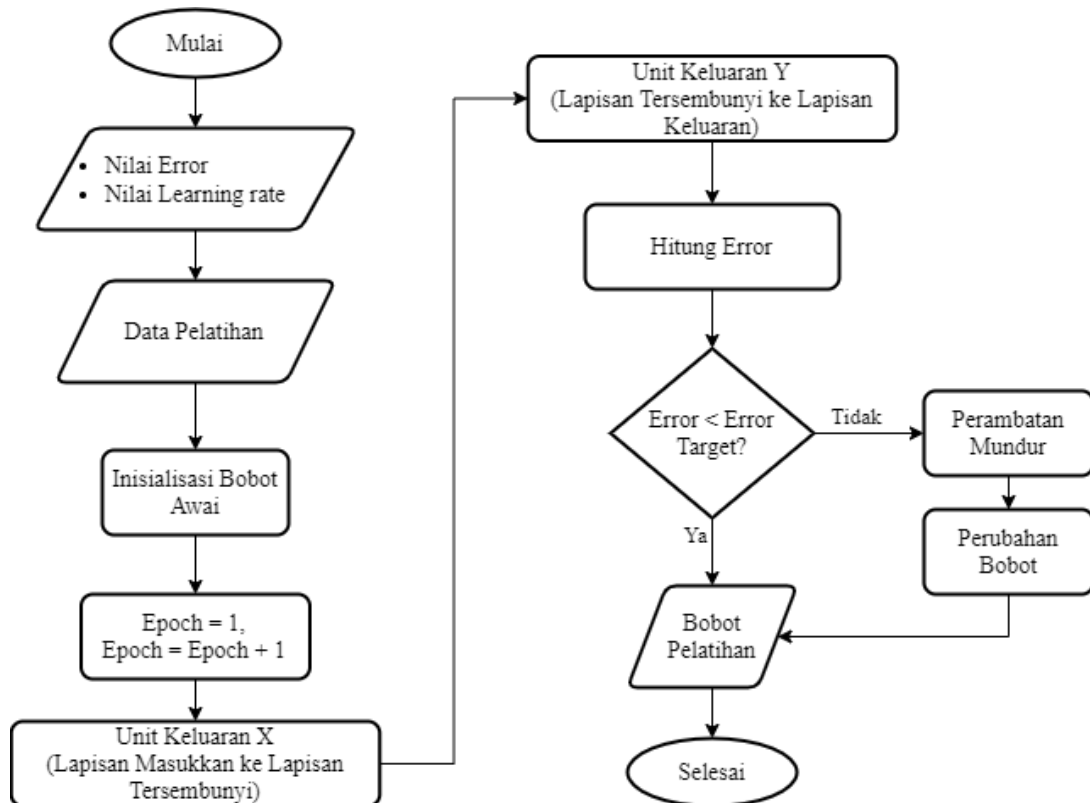


Gambar 5. Alur Pembobotan menggunakan TF-IDF

2.4 Pelatihan dan pengujian dengan backpropagation neural network.

Setelah diberikan bobot, selanjutnya dimasukkan ke dalam *Backpropagation Neural Network* (BNN) untuk dilakukan pelatihan. BNN ini digunakan untuk proses klasifikasi dua sentimen, yaitu sentimen positif dan negatif. BNN mempunyai 2 bagian yaitu *Propagation* dan *Weigh update*, dimana *propagation* terdiri dari forward dan backward [12][13]. Dengan dua *propagation* tersebut bisa mendapatkan hasil optimal dengan iterasi yang sedikit.

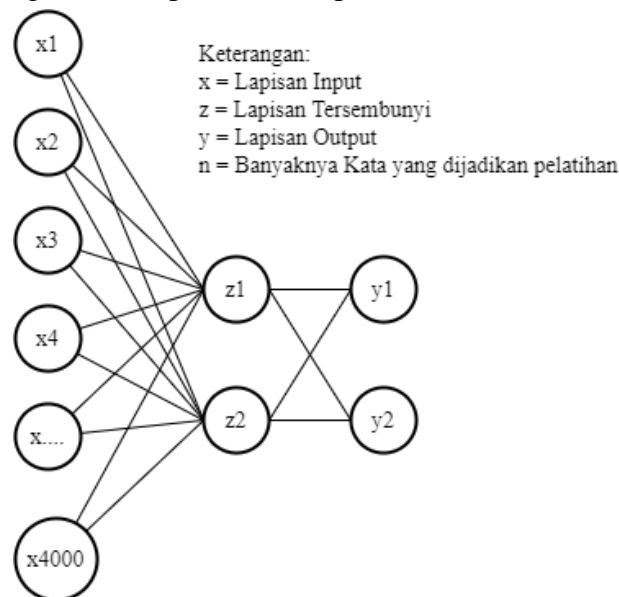
Setelah proses pelatihan selesai, selanjutnya dilakukan pengujian dengan data uji yang telah disiapkan. Alur pelatihan menggunakan BNN seperti terlihat pada Gambar 6.



Gambar 6. Alur Proses Pelatihan menggunakan BNN

3. Hasil dan Pembahasan

Arsitektur *Backpropagation Neural Network* yang digunakan pada penelitian ini terdiri dari 4000 *node input layer*, 2 *node hidden layer*, dan 2 *node output layer*. Arsitektur BNN tersebut memproses data input cukup lama dikarenakan hanya menggunakan 2 *hidden layer*, sehingga untuk mempercepat waktu pemodelan bisa ditambahkan *hidden layer*-nya. Arsitektur BNN yang digunakan seperti terlihat pada Gambar 7.



Gambar 7. Arsitektur BNN Analisis Sentimen

Penelitian ini menggunakan 158 *data training* yang berisi 4000 kata. Hasilnya jumlah klasifikasi benar dan nilai akurasi semakin bertambah hanya saja penambahannya cenderung tidak *significant* sehingga hanya digunakan 5 kali iterasi. Pembatasan ini terjadi dikarenakan spesifikasi hardware yang digunakan terbatas sehingga untuk mengantisipasi waktu yang lama dibatasi jumlah iterasi dan jumlah *node hidden layer*. Hasil perhitungan akurasi pelatihan seperti terlihat pada Tabel 1.

Tabel 1. Tabel Akurasi Pelatihan

Iterasi	Jumlah Data Training	Jumlah Klasifikasi Benar	Akurasi
1	158	128	81.0%
2	158	129	81.6%
3	158	134	84.8%
4	158	135	85.4%
5	158	136	86.0%

Setelah model analisis sentimen terbentuk dari hasil pelatihan, selanjutnya model tersebut diuji menggunakan data uji sebanyak 69 data. Hasil nilai akurasi menunjukkan angka 62.3%. Hasil perhitungan akurasi pengujian seperti terlihat pada Tabel 2.

Tabel 2. Tabel Akurasi Pengujian

		Hasil Prediksi	
		Positif	Negatif
Target	Positif	32	16
	Negatif	10	11

$$\text{Akurasi} = \frac{32 + 11}{32 + 11 + 16 + 10} * 100 = 62.3\%$$

4. Kesimpulan

Penelitian telah selesai dilakukan, menunjukkan bahwa sistem dapat mendeteksi sentimen pada Twitter dengan menggunakan metode *Backpropagation Neural Network* (BNN). Akurasi yang didapatkan dari hasil pengujian adalah sebesar 62.3% dengan 5 kali iterasi dan 2 *node hidden layer*. Perhitungan yang terjadi menunjukkan penambahan jumlah iterasi tidak sebanding dengan bertambahnya nilai akurasi. Penambahan nilai akurasi tidak terjadi secara *significant*. Itu terjadi karena keterbatasan spesifikasi *hardware* yang digunakan sehingga perlu dibatasi jumlah *node hidden layer* dan jumlah iterasi.

Sebagai saran pada penelitian selanjutnya, untuk mendapatkan nilai akurasi yang lebih tinggi, maka perlu menambahkan jumlah *node hidden layer*. Spesifikasi *hardware* komputer yang digunakan juga berpengaruh terhadap proses perhitungan tiap iterasi, sehingga untuk mendapatkan nilai akurasi yang bagus dan proses perhitungan yang cepat diperlukan spesifikasi hardware komputer yang tinggi.

Daftar Pustaka

- [1] S. Kemp, "Digital 2020: Indonesia," [dateportal.com](https://datareportal.com/reports/digital-2020-indonesia/), 2020. [Online]. Available:

- <https://datareportal.com/reports/digital-2020-indonesia>. [Accessed: 18-Nov-2020].
- [2] Fitriyyah, S. N. J., Safriadi, N., & Pratama, E. E. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 5(3), 279-285.
 - [3] Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng-Seri Informatika*, 3(1), 50-59.
 - [4] Nugroho, A. (2018). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstraksi Fitur N-Gram. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 2(2), 200-209.
 - [5] Assuja, M. A., & Saniati, S. (2016). Analisis Sentimen Tweet Menggunakan Backpropagation Neural Network. *Jurnal Teknoinfo*, 10(2), 48-53.
 - [6] Taufik, I., & Pamungkas, S. A. (2018). Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma Support Vector Machine (SVM). *LOGIK@*, 8(1), 69-79.
 - [7] Wibisono, G., & Hermawan, A. (2019). Faktor-Faktor Penentu Gejala Penyakit Kanker Payudara Dengan Pendekatan Jaringan Saraf Tiruan. *JASIEK (Jurnal Aplikasi Sains, Informasi, Elektronika dan Komputer)*, 1(1), 1-6.
 - [8] Fachrie, M., & Wibowo, A. P. (2018). Jaringan Syaraf Tiruan Untuk Memprediksi Kinerja Satpam. *JIKO (Jurnal Informatika dan Komputer)*, 3(1), 46-51.
 - [9] Hidayati, N., Fachrie, M., & Wibowo, A. P. (2019). Estimation of Time Voting in Elections Using Artificial Neural Network. *Compiler*, 8(2), 131-138.
 - [10] Waluyo, T., Hermawan, A., & Wibowo, A. P. (2019). Prediksi Penjualan Sepeda Motor Honda Menggunakan Jaringan Syaraf Tiruan. *Journal Of Information System Management*, 1(1), 31-35.
 - [11] Andayani, S., & Ryansyah, A. (2017). Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen. *JuSiTik: Jurnal Sistem dan Teknologi Informasi Komunikasi*, 1(1), 53-62.
 - [12] Katyal, R. (2015). Back Propagation Neural Network based Emotion Recognition System. *International Journal of Engineering Trends and Technology*, 22(4), 148–152. doi:10.14445/22315381/ijett-v22p231
 - [13] Rahayu, D., Wihandika, R. C., & Perdana, R. S. (2018). Implementasi Metode Backpropagation Untuk Klasifikasi Kenaikan Harga Minyak Kelapa Sawit. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X*.

