



Analisis Sentimen Masyarakat terhadap Teroris dalam Media Sosial Twitter menggunakan NLP

Shobrina Fathoniah¹, Chaerur Rozikin²

^{1,2}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

Received: 14 Juli 2022
Revised: 17 Juli 2022
Accepted: 19 Juli 2022

Abstract

The use of social media is not only to communicate between humans and other humans, but nowadays, social media can be a place to exchange ideas about various things, such as Twitter social media. Twitter is widely used in Indonesia to discuss various things that are currently happening or phenomenal events such as terrorism. So that an analysis of public sentiment against terrorism can be carried out on Twitter using NLP and LSTM. The use of NLP for preprocessing data is to remove urls and mentions and tokenization of tweet sentences. Meanwhile, the use of LSTM when modeling can provide an accuracy rate of 59.34% with a unit value of 20.

Keywords: Sentiment Analysis, Terrorism, Twitter, NLP

(*) Corresponding Author: shobrina.fathoniah18099@student.unsika.ac.id

How to Cite: Fathoniah, S., & Rozikin, C. (2022). Analisis Sentimen Masyarakat terhadap Teroris dalam Media Sosial Twitter menggunakan NLP. *Jurnal Ilmiah Wahana Pendidikan*, 8(13), 412-419. <https://doi.org/10.5281/zenodo.6962682>.

PENDAHULUAN

Penyampaian opini atau pemikiran masyarakat saat ini tidak hanya terjadi saat melakukan musyawarah bersama untuk mendapatkan penyelesaian dari masalah yang ada. Namun penyampaian opini saat ini telah meluas dan dapat dilakukan dimana saja dan kapan saja melalui media sosial, salah satunya yaitu Twitter. Pengguna Twitter di Indonesia per Januari 2022, mendapatkan urutan ke-5 pengguna Twitter terbanyak di dunia (Annur, 2022). Twitter sebagai tempat masyarakat dalam menyampaikan pendapat atau opininya mengenai berbagai hal, seperti pemerintahan, keadaan sekitar, ataupun fenomena yang tidak biasa terjadi di Indonesia, salah satunya adalah terorisme.

Terorisme menurut KBBI (Kamus Besar Bahasa Indonesia) adalah suatu cara yang menggunakan kekerasan dalam mencapai tujuannya sehingga menimbulkan ketakutan dalam masyarakat (Badan Pengembangan dan Pembinaan Bahasa, 2016). Terorisme di Indonesia telah terjadi sejak tahun 2000 (Harruma, 2022). Terdapat berbagai macam tindakan terorisme yang dilakukan, namun yang paling sering terjadi adalah peledakan bom, baik bom bunuh diri ataupun meledakan bom yang disimpan dalam sebuah gedung (Harruma, 2022). Tindakan terorisme yang baru ini terjadi di Indonesia adalah bom bunuh diri yang dilakukan di Gereja Katedral di Jalan Kajaolalido, Kota Makassar. Kejadian tersebut terjadi pada hari minggu tanggal 28 Maret 2021 (Yunus, 2021).

Kegiatan terorisme yang terjadi di Indonesia menimbulkan berbagai macam pendapat, penilaian, sentimen, dan sikap masyarakat dalam menanggapi. Pendapat dan penilaian masyarakat yang dituliskan pada kolom twitter sangat

banyak dan beragam sehingga sulit dilakukan analisis data pendapat dan penilaian masyarakat. Analisis data pendapat dan penilaian sangat penting untuk menilai kecenderungan atau sentimen masyarakat terhadap isu tertentu seperti isu terorisme.

Menurut penelitian yang dilakukan oleh Jashubhai mengenai *Opinion mining* pada berita utama surat kabar menggunakan SVM dan NLP yang menghasilkan akurasi 87.11% (Jashubhai Rameshbhai & Paulose, 2019). Penelitian yang dilakukan oleh mengenai penggunaan *Natural Language Processing* (NLP) sebagai tahap *pre-processing* sedangkan untuk pembuatan model dilakukan dengan menggunakan Naïve Bayes dalam analisis sentimen isu radikalisme menghasilkan akurasi sebesar 86% (Sundara et al., 2020). Menurut penelitian yang dilakukan oleh Liao dalam penggunaan CNN dan NLP untuk memahami sebuah situasi berdasarkan data dari twitter menghasilkan akurasi yaitu sebesar 75,39% (Liao et al., 2017). Penelitian yang dilakukan oleh Fairuz mengenai Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter dengan menggunakan NLP dan K-NN menghasilkan nilai akurasi sebesar 85% (Fairuz et al., 2021). Penelitian yang dilakukan oleh Novianto dan Munasatya dengan menggunakan objek penelitian yaitu tweet mengenai Presiden Jokowi menggunakan *Natural Language Processing* (NLP), dalam tahap *preprocessing* data tweet dapat memudahkan dalam pengenalan tweet sehingga dapat menghasilkan tingkat akurasi yaitu sebesar 93,26% (Munasatya & Novianto, 2020). Penelitian tersebut menggunakan MLP dalam pembuatan modelnya. Namun dalam penelitian mengenai perbandingan penggunaan *deep learning* saat melakukan analisis sentimen, antara MLP dan LSTM, LSTM memberikan hasil performa yang lebih baik daripada saat menggunakan MLP (Dashtipour et al., 2021).

Berdasarkan permasalahan tersebut, tujuan penelitian ini adalah membuat model sentimen analisis menggunakan model LSTM. Algoritma LSTM adalah salah satu jenis dari *machine learning* dan termasuk ke dalam bagian RNN atau *Recurrent Neural Network* (Wisyaludin et al., 2020). Arsitektur dari RNN dapat mempengaruhi nilai *output* yang sebelumnya pada perhitungan *output* saat ini (Wisyaludin et al., 2020). LSTM biasanya digunakan dalam memprediksi urutan (*sequence prediction*) (Navin Kumar Manaswi, 2018).

Tahapan penelitian ini yaitu dataset diambil dari kaggle kemudian dilakukan *preprocessing* menggunakan NLP. *Preprocessing data* adalah tahap dimana data asli yang belum diubah dilakukan sebuah proses yang bertujuan agar data dapat digunakan sebelum data tersebut digunakan untuk menyesuaikan dan mengevaluasi model *machine learning* (Brownlee, 2020). NLP adalah salah satu bidang dari kecerdasan buatan yang mencoba untuk memproses dan menganalisis mengenai data bahasa alami (Vasiliev, 2020).

Hasil dari *preprocessing* kemudian dilakukan pengklasifikasian berdasarkan kelas negatif, positif dan netral. Klasifikasi adalah permasalahan yang meliputi dengan mengambil masukan dan menentukan kelas N manakah milik mereka (Marsland, 2014). Sedangkan model *machine learning* adalah sebuah *file* yang telah dilatih untuk mengenali tipe dari sebuah pola (Radich & Cowley, 2021). Sehingga, klasifikasi model adalah menentukan kelas N sebuah model agar dapat mengenali pola yang diberikan.

Model yang telah dibuat kemudian dilakukan evaluasi dengan memberikan input pada model menggunakan data validasi yang telah disiapkan sebelumnya lalu

menghitung nilai akurasi, *precision*, dan *recall*. Akurasi adalah persentase kerentanan yang telah teridentifikasi benar baik sebagai bernilai benar positif atau benar negatif (Cheirdari & Karabatis, 2017). *Precision* adalah persentase kerentanan yang diidentifikasi sebagai bernilai benar positif (Cheirdari & Karabatis, 2017). Sedangkan *recall* adalah persentase dari kerentanan yang bernilai benar positif yang telah teridentifikasi (Cheirdari & Karabatis, 2017).

Penelitian yang dilakukan menghasilkan tingkat akurasi tertinggi yaitu 59.34% dengan menggunakan nilai *units* dalam LSTM yaitu sebesar 20. *Units* sendiri adalah salah satu parameter dalam penggunaan algoritma LSTM dengan *library* Keras yang dituliskan dalam bentuk bilangan bulat positif dan dimaksudkan untuk menentukan jumlah sel memori (Jason Brownlee, 2017). Sehingga dalam penelitian ini LSTM dan NLP masih belum dapat bekerja dengan baik berdasarkan dari nilai akurasi, *precision*, dan *recall* dalam mengklasifikasi jenis dari teks yang diberikan baik bernilai positif, negatif, maupun netral. Hal tersebut dikarenakan oleh masih perlu dilakukan penambahan tahap *preprocessing* dataset agar dapat memberikan hasil yang lebih baik.

METODOLOGI PENELITIAN

Dataset yang digunakan dalam penelitian ini bersumber dari Kaggle milik M Razif Rizqullah, dengan judul dataset yaitu Indonesian Tweet About Terrorism (Teroris). Dataset tersebut dicari menggunakan kata kunci pencarian yaitu teroris, berisikan data tweet asli, *reply*, dan *quote tweet* (Rizqullah, 2022). Berikut ini adalah daftar parameter dari dataset tersebut.

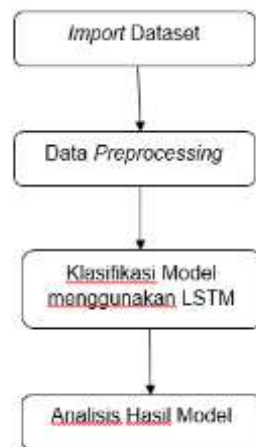
Tabel 1. Daftar Parameter dalam Dataset

Parameter	Keterangan
<i>tweet id</i>	Id dari <i>tweet</i>
<i>name</i>	Nama penulis <i>tweet</i>
<i>username</i>	Username penulis <i>tweet</i>
<i>tweet text</i>	Text dari <i>tweet</i> yang dituliskan
<i>reference type</i>	Jenis kegiatan yang dilakukan seperti, <i>original</i> atau <i>reply to</i>
<i>reference id</i>	Id dari jenis kegiatan yang dilakukan
<i>created at</i>	Tanggal penulisan <i>tweet</i>
<i>quote</i>	Jumlah yang melakukan <i>quote</i> terhadap tweetnya
<i>like</i>	Jumlah yang melakukan <i>like</i> terhadap tweetnya
<i>reply</i>	Jumlah yang melakukan <i>reply</i> terhadap tweetnya

Sumber: Rizqullah (2022)

Dalam penelitian ini yang digunakan hanyalah parameter *tweet text* dan selanjutnya akan dilakukan pemberian label oleh penulis untuk menentukan apakah tweet tersebut berisikan tweet yang bernilai positif atau negatif ataupun netral.

Penelitian ini akan dilakukan dengan menggunakan tahap *preprocessing* yaitu NLP (*Natural Language Processing*) dan pembuatan model ML dengan menggunakan LSTM (*Long Short Term Memory Network*). Adapun tahapan penelitian yang dilakukan adalah sebagai berikut.



Gambar 1. Tahapan Penelitian

Import Dataset

Import dataset digunakan untuk membaca *file* yang berasal dari luar. Penulis akan memasukkan dataset *Indonesian Tweet About Terrorism* (Teroris) sebagai langkah awal dalam mengolah data. Saat *import dataset* dapat dilakukan dengan meng-*upload file* dataset ke dalam Google Colab.

Data Preprocessing

Pada tahap *data preprocessing* dalam penelitian ini, penulis akan menggunakan NLP (*Natural Language Processing*). Penulis akan melakukan seleksi terhadap data tweet, penghapusan mention, tanda baca, *link* gambar twitter, dan *link* website yang dituliskan dalam tweet. Selanjutnya, penulis akan mengelompokkan data tweet menjadi 1 untuk data tweet yang bernilai positif, 2 untuk data tweet yang bernilai negatif, dan 0 untuk data tweet yang bernilai netral. Terakhir, dilakukan Tokenization. Tokenization digunakan untuk mengubah teks menjadi token, teks tersebut dapat berupa kata, nomor, atau tanda baca (Vasiliev, 2020).

Klasifikasi Model menggunakan LSTM

Pada tahap pengklasifikasian model, model yang diusulkan akan diuji untuk melihat hasil akurasi. Percobaan yang dilakukan pada penelitian ini menggunakan algoritma LSTM (*Long Short Term Memory*). Penggunaan algoritma LSTM dikarenakan berdasarkan penelitian yang dilakukan Dashtipour et al. (2021), LSTM menghasilkan performa yang lebih baik dibandingkan penggunaan algoritma MLP.

Analisis Hasil Model

Setelah dilakukan pemodelan, maka akan dilanjutkan dengan proses analisis model klasifikasi untuk melihat akurasi model-model yang telah dibuat. Pada tahap ini, akan dilakukan perbandingan nilai akurasi, *precision*, dan *recall* antar masing-masing model yang diterapkan.

HASIL & PEMBAHASAN

Hasil

Pada penelitian ini, peneliti menggunakan situs Google Colab dalam melakukan analisis sentimen terhadap terorisme di Indonesia dengan bahasa pemrograman Python.

Import Dataset

Saat *import* dataset, kami *upload* dataset ke dalam Google Drive sehingga saat penggunaan di Google Colab hanya menuliskan lokasi *file* dataset yang telah di-*upload* tersebut, seperti berikut ini.

```
import os
import pandas as pd

df = pd.read_csv('/content/drive/MyDrive/Jurnal/Dataset/teroris.csv')
df_origin = df
```

Gambar 2. Source Code Import Dataset

Berikut ini adalah data awal dari dataset yang diambil dari Kaggle milik M Razif Rizqullah.

	tweet_id	name	username	tweet_text	reference_type	reference_id	created_at	like	quote	reply	retweet
3601	1502924699332153348	davinaputi	davinaputi21	Mengenal Lebih Jauh Sosok Komika Jebolan SUKA...	original	NaN	2022-09-13T08:29:05.000Z	0	0	0	0
4661	1502884342028443653	monakusa	monakusa	Apa yg telah dilakukan aparat penegak hukum su...	original	NaN	2022-09-13T05:48:44.000Z	0	0	0	0
9540	1502665163640445952	Sukio	hayomarr	@dixpinh Kenapa hanya Islam yg dilabeli teror...	replied_to	1.502470e+18	2022-09-12T15:17:48.000Z	0	0	0	0
8072	1502817364264835070	Sambang	Sambang45263409	@KutahnaWangsa @P3n3l4j4n_id "Teroris" ini ha...	quoted	1.502816e+18	2022-09-13T01:22:11.000Z	0	0	0	1

Gambar 3. Data Awal Dataset

Data Preprocessing

Pada tahap ini, dilakukan perbaikan terhadap dataset yang telah di-*upload* dengan menggunakan NLP. Berikut ini tahapan yang dilakukan yaitu:

1. Seleksi data *tweet*

Penyeleksian data *tweet* dilakukan untuk menghilangkan data *tweet* yang memiliki teks yang sama persis. Saat penyeleksian ini, menggunakan aplikasi Microsoft Excel dan menghapus semua data yang teridentifikasi duplikat.

2. Data *Cleaning*

Data *cleaning* dilakukan untuk membersihkan data dari hal yang tidak diperlukan. Dalam dataset asli, terdapat banyak parameter yang tidak diperlukan. Oleh karena itu, kolom-kolom tersebut dihapuskan dan tersisa hanyalah kolom *tweet text*. Selanjutnya *tweet text* tersebut diperbaiki kembali dengan menghapuskan *mention* akun lain, tanda baca yang digunakan, *link* gambar yang disediakan twitter, dan *link website* yang dituliskan dalam *tweet*.

3. *Labeling*

Pemberian label pada setiap teks *tweet* dilakukan untuk mengetahui *tweet* tersebut termasuk ke dalam jenis teks yang bernilai positif, negatif, atau netral. Pemberian nilai dinotasikan dengan angka. Dimana angka 1 untuk teks bernilai positif, angka 2 untuk teks bernilai negatif, dan angka 0 untuk teks bernilai netral.

Sehingga jumlah teks *tweet* yang bernilai positif yaitu 324 *tweet*, bernilai negatif yaitu 537 *tweet*, dan bernilai netral yaitu 209 *tweet*.

4. Tokenization

Tokenisasi dilakukan untuk mengubah teks baik kata, angka, ataupun tanda baca menjadi token.

Klasifikasi Model menggunakan LSTM

Pada tahap ini, dilakukan pembuatan model yang dilakukan dengan menggunakan algoritma LSTM (*Long Short Term Memory*). Dalam pembuatan model, dataset yang telah diperbaiki sebelumnya, dibagi menjadi tiga yaitu data *training*, data *validation*, dan data *testing*. Sehingga dalam penelitian ini, menggunakan data *training* berjumlah 749 *tweet*, data *validation* berjumlah 214 *tweet*, dan data *testing* berjumlah 107 *tweet*.

Analisis Hasil Model

Setelah model dibuat dengan menggunakan algoritma LSTM, dilakukan pengujian terhadap model tersebut dengan menggunakan data validasi. Berikut ini adalah hasil akurasi, *precision*, dan *recall* dari model yang telah dibuat.

Tabel 2. Hasil Akurasi, *Precision*, dan *Recall* Model

Units	Akurasi (%)	<i>Precision</i> (%)	<i>Recall</i> (%)
10	57.00	53.38	52.94
20	59.34	56.56	55.98
30	57.47	54.83	54.62
40	57.94	56.55	56.20
50	54.20	51.27	53.54
60	58.87	56.91	55.15
70	52.80	50.54	51.15
80	52.80	50.40	50.15
90	50.93	49.92	47.83
100	53.27	52.52	49.17

Berdasarkan penelitian yang dilakukan, akurasi tertinggi terjadi saat jumlah *units* yang digunakan dalam model LSTM yaitu 20 dengan persentase akurasi sebesar 59.34%.

KESIMPULAN

Berdasarkan pengujian yang telah dilakukan dengan tujuan untuk mengetahui pengaruh dari penggunaan LSTM (*Long Short Term Memory*) dan NLP (*Natural Language Processing*) dalam menganalisis sentimen masyarakat terhadap terorisme di media sosial twitter. Penelitian ini memberikan hasil bahwa dengan menggunakan LSTM dan memberikan nilai *units* sebesar 20 akan menghasilkan tingkat akurasi sebesar 59.34%. Dimana tingkat akurasi tersebut merupakan tingkat tertinggi dibandingkan saat penggunaan nilai unit sebesar 10, 30, 40, 50, 60, 70, 80, 90, dan 100. Penelitian yang dilakukan masih memerlukan beberapa pengembangan agar dapat menghasilkan model dengan tingkat akurasi yang lebih tinggi. Hal ini dapat dicapai dengan penambahan jumlah dataset yang digunakan, penggunaan

algoritma klasifikasi lain yang lebih tepat, dan penambahan metode saat melakukan *preparation data*.

DAFTAR PUSTAKA

- Annur, C. M. (2022, March 23). *Pengguna Twitter Indonesia Masuk Daftar Terbanyak di Dunia, Urutan Berapa?* | Databoks. <https://databoks.katadata.co.id/datapublish/2022/03/23/pengguna-twitter-indonesia-masuk-daftar-terbanyak-di-dunia-urutan-berapa>
- Badan Pengembangan dan Pembinaan Bahasa. (2016). *Hasil Pencarian - KBBI Daring*. <https://kbbi.kemdikbud.go.id/entri/terorisme>
- Brownlee, J. (2020, June 30). *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python*. Machine Learning Mastery. https://books.google.co.id/books?hl=id&lr=&id=uAPuDwAAQBAJ&oi=fnd&pg=PP1&dq=data+preparation&ots=Cl3Mzi9Op_&sig=g9QrxQ-gtKS5CYSPaH0eGWRyh-U&redir_esc=y#v=onepage&q=data+preparation&f=false
- Cheirdari, F., & Karabatis, G. (2017). On the Verification of Software Vulnerabilities During Static Code Analysis Using Data Mining Techniques. In Adrian Paschke, Christophe Debruyne, Claudio Agostino Ardagna, Hervé Panetto, Mike Papazoglou, Robert Meersman, & Walid Gaaloul (Eds.), *Springer International Publishing* (1st ed.). https://www.google.co.id/books/edition/On_the_Move_to_Meaningful_Internet_Systems/4_06DwAAQBAJ?hl=id&gbpv=0
- Dashtipour, K., Gogate, M., Adeel, A., Larijani, H., & Hussain, A. (2021). Sentiment analysis of persian movie reviews using deep learning. *Entropy*, 23(5), 1–16. <https://doi.org/10.3390/e23050596>
- Fairuz, A. L., Ramadhani, R. D., & Tanjung, N. A. F. (2021). Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter. *Journal of Dinda : Data Science, Information Technology, and Data Analytics*, 1(1), 42–51. <https://doi.org/10.20895/dinda.v1i1.180>
- Harruma, I. (2022, April 28). *Kasus-Kasus Terorisme di Indonesia dan Penyelesaiannya Halaman all - Kompas.com*. <https://nasional.kompas.com/read/2022/04/28/01150071/kasus-kasus-terorisme-di-indonesia-dan-penyelesaiannya?page=all>
- Jashubhai Rameshbhai, C., & Paulose, J. (2019). Opinion mining on newspaper headlines using SVM and NLP. *International Journal of Electrical and Computer Engineering (IJECE)*, 9(3), 2152–2163. <https://doi.org/10.11591/ijece.v9i3.pp2152-2163>
- Jason Brownlee. (2017). *Long Short-Term Memory Networks With Python - Google Books* (1.5). Machine Learning Mastery. https://www.google.co.id/books/edition/Long_Short_Term_Memory_Networks_With_Pyt/m7SoDwAAQBAJ?hl=id&gbpv=0
- Liao, S., Wang, J., Yu, R., Sato, K., & Cheng, Z. (2017). CNN for situations understanding based on sentiment analysis of twitter data. *Procedia Computer Science*, 111. <https://doi.org/10.1016/j.procs.2017.06.037>
- Marsland, S. (2014). *MACHINE LEARNING An Algorithmic Perspective Second*

- Edition* (R. Herbrich & T. Graepel (eds.); Second Edition). CRC Press.
[https://doc.lagout.org/science/Artificial Intelligence/Machine learning/Machine Learning_ An Algorithmic Perspective %282nd ed.%29 %5BMarsland 2014-10-08%5D.pdf](https://doc.lagout.org/science/Artificial%20Intelligence/Machine%20learning/Machine%20Learning%20An%20Algorithmic%20Perspective%202nd%20ed.%20%5BMarsland%202014-10-08%5D.pdf)
- Munasatya, N., & Novianto, S. (2020). Natural Language Processing untuk Sentimen Analisis Presiden Jokowi Menggunakan Multi Layer Perceptron. *Techno.Com*, 19(3), 237–244. <https://doi.org/10.33633/tc.v19i3.3630>
- Navin Kumar Manaswi. (2018). *Deep Learning with Applications Using Python* (1st ed.). Apress. https://www.google.co.id/books/edition/Deep_Learning_with_Applications_Using_Py/5HNUdWAAQBAJ?hl=id&gbpv=0
- Radich, Q., & Cowley, E. (2021, May 26). *What is a machine learning model?* | Microsoft Docs. <https://docs.microsoft.com/en-us/windows/ai/windows-ml/what-is-a-machine-learning-model>
- Rizqullah, M. R. (2022, March 15). *Indonesian Tweet About Terrorism (Teroris)* | Kaggle. <https://www.kaggle.com/datasets/linkgish/indonesian-tweet-about-teroris-terrorism>
- Sundara, T. A., Arnas, S. E., & Sotar. (2020). Naïve Bayes Classifier untuk Analisis Sentimen Isu Radikalisme. *SISFOTEK*, 4. <http://www.seminar.iaii.or.id/index.php/SISFOTEK/article/view/159/141>
- Vasiliev, Y. (2020). *NLP with Python & SpaCy* (1st ed., Issue 1). No Starch Press. https://www.google.co.id/books/edition/Natural_Language_Processing_with_Python/IVv6DwAAQBAJ?hl=id&gbpv=0
- Wisyaladin, M. K., Luciana, G. M., & Pariaman, H. (2020). Pendekatan Long Short-Term Memory untuk Memprediksi Kondisi Motor 10 kV pada PLTU Batubara. *Jurnal Kilat*, 9(2), 311–318.
- Yunus, M. (2021, March 29). *Data Serangan Teroris di Indonesia : Rata-rata Lebih 2 Kali Setiap Bulan - Suara Sulsei*. <https://sulsei.suara.com/read/2021/03/29/053004/data-serangan-teroris-di-indonesia-rata-rata-lebih-2-kali-setiap-bulan?page=all>