

**ANALISIS PERBANDINGAN LABELING DARI LEXICON BASED DAN  
LIBRARY TEXTBLOB PADA ALGORITMA LONG-SHORT TERM  
MEMORY TERHADAP ANALISIS SENTIMEN KOMENTAR YOUTUBE  
MENGENAI PENERAPAN 5G DI INDONESIA**

**SKRIPSI**

diajukan untuk memenuhi sebagai syarat untuk memperoleh gelar Sarjana Teknik  
pada Program Studi Sistem Telekomunikasi



oleh

Prihantoro Tri Nugroho

NIM 1905856

**PROGRAM STUDI SISTEM TELEKOMUNIKASI  
KAMPUS UPI DI PURWAKARTA  
UNIVERSITAS PENDIDIKAN INDONESIA  
2023**

**ANALISIS PERBANDINGAN LABELING DARI LEXICON BASED DAN  
LIBRARY TEXTBLOB PADA ALGORITMA LONG-SHORT TERM  
MEMORY TERHADAP ANALISIS SENTIMEN KOMENTAR YOUTUBE  
MENGENAI PENERAPAN 5G DI INDONESIA**

Oleh

Prihantoro Tri Nugroho

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar  
Sarjana Teknik pada Program Studi Sistem Telekomunikasi

© Prihantoro Tri Nugroho 2023

Universitas Pendidikan Indonesia

Agustus 2023

Hak Cipta dilindungi undang-undang.

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,  
dengan dicetak ulang, difoto kopi, atau cara lainnya tanpa ijin dari penulis.

**LEMBAR PENGESAHAN**

**PRIHANTORO TRI NUGROHO**

**ANALISIS PERBANDINGAN LABELING DARI LEXICON BASED DAN  
LIBRARY TEXTBLOB PADA ALGORITMA LONG-SHORT TERM  
MEMORY TERHADAP ANALISIS SENTIMEN KOMENTAR YOUTUBE  
MENGENAI PENERAPAN 5G DI INDONESIA**

disetujui dan disahkan oleh pembimbing:

**Pembimbing I**



**Hafivyan Putra Pratama, S.ST., M.T.**

**NIP. 920190219921224101**

**Pembimbing II**

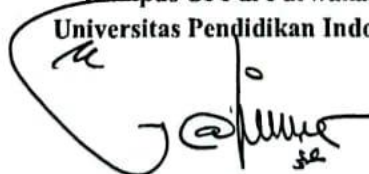


**Ahmad Fauzi, S.Si., M.T.**

**NIP. 920171219820915101**

**Mengetahui,**

**Ketua Program Studi Sistem Telekomunikasi  
Kampus UPI di Purwakarta  
Universitas Pendidikan Indonesia**



**Galura Muhammad Suranegara, S.Pd., M.T.**

**NIP. 920190219920111101**

## **PERNYATAAN KEASLIAN SKRIPSI DAN BEBAS PLAGIARISME**

Dengan ini saya menyatakan bahwa skripsi dengan judul "Analisis Perbandingan *Labeling* dari *Lexicon Based* dan *Library Textblob* pada Algoritma *Long-Short Term Memory* Terhadap Analisis Sentimen Komentar Youtube Mengenai Penerapan 5G di Indonesia" ini beserta seluruh isinya adalah benar-benar karya saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya saya ini.

Purwakarta, Agustus 2023

Penulis,

Prihantoro Tri Nugroho

NIM. 1905856

## KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan Rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan lancar.

Penyusunan skripsi yang berjudul “Analisis Perbandingan *Labeling* dari *Lexicon Based* dan *Library Textblob* Pada Algoritma *Long-Short Term Memory* Terhadap Analisis Sentimen Komentar Youtube Mengenai Penerapan 5G di Indonesia” ini bertujuan untuk memenuhi salah satu syarat dalam memperoleh gelar Sarjana Teknik pada Program Studi Sistem Telekomunikasi Universitas Pendidikan Indonesia Kampus Purwakarta.

Dalam penulisan skripsi ini penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan karena kekurangan dan keterbatasan yang penulis miliki. Maka dari itu saran dan masukan yang membangun sangat dibutuhkan demi perbaikan skripsi ini selanjutnya menjadi lebih baik.

Penulis sangat berterima kasih terhadap semua pihak yang telah membantu dan terlibat dalam proses penulisan skripsi ini. Harapan penulis semoga skripsi ini dapat memberikan manfaat khususnya kepada penulis, dan umumnya bagi pembaca serta berguna bagi proses pendidikan negeri ini.

Purwakarta, Agustus 2023

Penulis,

Prihantoro Tri Nugroho

NIM. 1905856

## UCAPAN TERIMAKASIH

Penyusunan skripsi ini tidak akan terselesaikan tanpa bantuan dari berbagai pihak. Pada kesempatan dengan segenap kerendahan hati penulis ingin menyampaikan ucapan terimakasih kepada:

1. Allah SWT yang telah memberikan nikmat, rahmat, dan karunia-Nya kepada penulis untuk menyelesaikan skripsi ini.
2. Bapak Hafiyyan Putra Pratama, S.ST., M.T. selaku Dosen Pembimbing 1 yang telah memberikan banyak motivasi, saran, dan masukan hingga skripsi ini dapat terselesaikan.
3. Bapak Ahmad Fauzi, S.Si., M.T. selaku Dosen Pembimbing 2 yang telah memberikan banyak motivasi, saran, dan masukan hingga skripsi ini dapat terselesaikan.
4. Seluruh Dosen Program Studi Sistem Telekomunikasi Universitas Pendidikan Indonesia yang telah memberikan ilmu pengetahuan yang bermanfaat selama saya mengikuti pendidikan di Sistem Telekomunikasi.
5. Kepada orang tua, kakak, dan adik penulis yang senantiasa memberikan doa, dukungan, dan motivasi kepada penulis. Terima kasih banyak.
6. Rekan-rekan mahasiswa Program Studi Sistem Telekomunikasi angkatan 2019 yang telah menemani dan memberikan dukungan selama kuliah.
7. Rekan-rekan Tim Magang *Data Analyst* Hukumonline yang telah menemani, membimbing, serta memberikan saran dan masukan selama proses magang.

**ANALISIS PERBANDINGAN LABELING DARI LEXICON BASED DAN  
LIBRARY TEXTBLOB PADA ALGORITMA LONG-SHORT TERM  
MEMORY TERHADAP ANALISIS SENTIMEN KOMENTAR YOUTUBE  
MENGENAI PENERAPAN 5G DI INDONESIA**

Prihantoro Tri Nugroho

NIM 1905856

**ABSTRAK**

Media sosial seperti Youtube merupakan media yang seringkali dipakai oleh masyarakat untuk mengemukakan opini serta pendapat mereka mengenai berbagai hal, salah satunya ialah perkembangan 5G di Indonesia. Masyarakat dapat menuliskan opini mereka melalui salah satu fitur yang ada di Youtube yaitu kolom komentar. Namun, semakin banyak dan beragam isi komentar maka semakin sulit apabila ingin menarik kesimpulan dari opini-opini tersebut, maka diperlukan sebuah sistem yang dapat melakukan hal itu, yaitu analisis sentimen. Analisis sentimen adalah sebuah sistem yang dapat menganalisa dokumen teks, lalu mengelompokkannya ke dalam beberapa sentimen seperti positif atau negatif. Dengan melakukan analisis sentimen, masyarakat akan mendapatkan informasi-informasi yang baru dan dapat diolah agar menghasilkan informasi yang bermanfaat. Penelitian ini terdiri dari beberapa tahapan sebelum menghasilkan sebuah sistem yang dapat menganalisis sentimen komentar Youtube yang berhubungan dengan perkembangan 5G di Indonesia. Tahap-tahap yang dilakukan adalah *scraping data* komentar, lalu *preprocessing* dimulai dari *case folding*, *cleansing*, normalisasi, dan *filtering/stopword*. Selanjutnya dilanjutkan proses *labeling* menggunakan dua metode yaitu *Library Textblob* dan *Lexicon Based*. Kemudian data di *split* dan di klasifikasikan menggunakan algoritma *Long-Short Term Memory* (LSTM). Hasil akhir dari kedua *labeling* didapatkan sentimen negatif yang lebih dominan dibandingkan dengan sentimen positif serta didapatkan nilai akurasi menggunakan *labeling* dari *lexicon* menghasilkan nilai akurasi yang lebih besar yaitu 86,4% dibandingkan dengan *labeling* yang menggunakan *library textblob* yang menghasilkan nilai akurasi sebesar 79,5%.

**Kata Kunci :** Analisis Sentimen, LSTM, *Lexicon Based*, *Library Textblob*.

**COMPARATIVE ANALYSIS OF LABELING FROM LEXICON BASED AND  
LIBRARY TEXTBLOB ON LONG-SHORT TERM MEMORY ALGORITHM  
TO ANALYSIS OF YOUTUBE COMMENTS SENTIMENT REGARDING 5G  
IMPLEMENTATION IN INDONESIA**

Prihantoro Tri Nugroho

NIM 1905856

**ABSTRACT**

*Social media such as Youtube is the media that is often used by the public to express their opinions and opinions on various matters, one of which is the development of 5G in Indonesia. The public can write their opinions through one of the features on Youtube, namely the comments column. However, the more and more varied the contents of comments, the more difficult it is to draw conclusions, so a system is needed that can do that, namely sentiment analysis. Sentiment analysis is a system that can analyze text documents, then group them into several sentiments such as positive or negative. By conducting sentiment analysis, the public will get new information and can be processed to produce useful information. This research consists of several stages before producing a system that can analyze the sentiment of Youtube comments related to the development of 5G in Indonesia. The steps taken are scraping comment data, preprocessing starting from case folding, cleansing, normalization, and filtering/sopword. Then proceed with the labeling process using two methods, namely Textblob Library and Lexicon Based. Then the data is split and classified using the Long-Short Term Memory (LSTM) algorithm. The final result of both labeling is negative sentiment which is more dominant than positive sentiment and the accuracy value obtained using labeling from lexicon produces a greater accuracy value of 86.4% compared to labeling using the textblob library which produces an accuracy value of 79.5%.*

**Keywords :** *Sentiment Analysis, LSTM, Lexicon Based, Library Textblob*



## DAFTAR ISI

LEMBAR HAK CIPTA .....	ii
LEMBAR PENGESAHAN .....	iii
PERNYATAAN KEASLIAN SKRIPSI DAN BEBAS PLAGIARISME .....	iv
KATA PENGANTAR .....	v
UCAPAN TERIMAKASIH.....	vi
ABSTRAK .....	vii
ABSTRACT.....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.4.1. Manfaat Teoritis .....	3
1.4.2. Manfaat Praktis .....	3
1.5 Batasan Masalah.....	4
1.6 Sistematika Penulisan.....	4
BAB II KAJIAN PUSTAKA .....	6
2.1 <i>Natural Language Processing (NLP)</i> .....	6
2.2 Penelitian Analisis Sentimen.....	6
2.2.1 <i>Scraping</i> .....	7

2.2.2	<i>Preprocessing</i> .....	7
2.2.2.1	<i>Case Folding</i> .....	8
2.2.2.2	<i>Cleansing</i> .....	8
2.2.2.3	<i>Normalisasi</i> .....	9
2.2.2.4	<i>Filtering / Stopword Removal</i> .....	9
2.2.3	<i>Labeling</i> .....	9
2.2.3.1	<i>Lexicon Based</i> .....	10
2.2.3.2	<i>Library Textblob</i> .....	10
2.2.4	<i>Modeling</i> .....	11
2.2.4.1	<i>Keras</i> .....	11
2.2.4.2	<i>Algoritma LSTM</i> .....	11
2.2.5	<i>Evaluasi Hasil</i> .....	12
2.2.5.1	<i>Classification Report</i> .....	12
2.2.5.2	<i>Confusion Matrix</i> .....	12
2.3	<i>Youtube API</i> .....	13
2.4	<i>Kinerja Model</i> .....	14
2.4.1	<i>Overfitting</i> .....	14
2.4.2	<i>Underfitting</i> .....	15
2.5	<i>Penelitian yang Relevan</i> .....	16
<b>BAB III METODE PENELITIAN</b> .....		18
3.1	<i>Studi Literatur</i> .....	18
3.2	<i>Analisis Kebutuhan</i> .....	19
3.3	<i>Perancangan Sistem</i> .....	19
3.4	<i>Implementasi</i> .....	20
3.5	<i>Pengujian dan Evaluasi Sistem</i> .....	21

BAB IV HASIL DAN PEMBAHASAN .....	22
4.1 Spesifikasi Kebutuhan Perangkat .....	22
4.2 Hasil Penelitian.....	22
4.2.1 <i>Scraping Data</i> .....	22
4.2.2 <i>Preprocessing</i> .....	23
4.2.3 <i>Labeling</i> .....	24
4.2.4 <i>Data Splitting</i> .....	27
4.2.5 <i>Modeling</i> dengan LSTM .....	27
4.2.6 Evaluasi Hasil.....	30
4.3 Analisis Hasil Penelitian .....	31
4.3.1 Analisa Hasil Pengujian .....	31
4.3.2 Perbandingan Hasil dengan 2 Metode <i>Labeling</i> .....	34
BAB V PENUTUP.....	35
5.1 Kesimpulan.....	35
5.2 Saran.....	36
DAFTAR PUSTAKA .....	37
LAMPIRAN-LAMPIRAN.....	40

## DAFTAR GAMBAR

Gambar 2. 1 Tahap <i>Preprocessing</i> .....	8
Gambar 2. 2 Tampilan Youtube API .....	14
Gambar 3. 1 Langkah-Langkah Penyelesaian Metode .....	18
Gambar 3. 2 Flowchart Penelitian.....	19
Gambar 4. 1 <i>Dataset Hasil Scraping</i> .....	23
Gambar 4. 2 Hasil proses <i>preprocessing</i> .....	24
Gambar 4. 3 Kamus <i>lexicon</i> .....	24
Gambar 4. 4 Presentase sentimen metode <i>lexicon based</i> .....	25
Gambar 4. 5 Hasil <i>scoring data</i> metode <i>lexicon based</i> .....	25
Gambar 4. 6 Presentase sentimen menggunakan <i>library textblob</i> .....	26
Gambar 4. 7 Hasil <i>polarity data</i> menggunakan <i>library textblob</i> .....	26
Gambar 4. 8 Model akurasi untuk metode <i>lexicon</i> dan <i>textblob</i> .....	29
Gambar 4. 9 Model <i>loss</i> untuk metode <i>lexicon</i> dan <i>textblob</i> .....	29
Gambar 4. 10 <i>Confusion matrix lexicon based</i> .....	30
Gambar 4. 11 <i>Confusion matrix library textblob</i> .....	31

## DAFTAR TABEL

Tabel 2. 1 Contoh Kamus Kata pada Metode <i>Lexicon Based</i> .....	10
Tabel 2. 2 <i>Confusion Matrix</i> .....	13
Tabel 2. 3 Perbandingan penelitian penulis dengan penelitian sebelumnya.....	16
Tabel 4. 1 Tahapan <i>Preprocessing</i> .....	23
Tabel 4. 2 <i>Classification report lexicon based</i> .....	30
Tabel 4. 3 <i>Classification report library textblob</i> .....	31
Tabel 4. 4 Perbandingan hasil dari metode <i>lexicon</i> dan <i>textblob</i> .....	34

## DAFTAR LAMPIRAN

Lampiran 1. <i>Source Code Scraping</i> .....	40
Lampiran 2. <i>Source Code Preprocessing</i> .....	41
Lampiran 3. <i>Source Code Labeling</i> .....	42
Lampiran 4. <i>Source Code Data Split</i> .....	43
Lampiran 5. <i>Source Code Modeling</i> .....	44
Lampiran 6. <i>Source Code Evaluasi</i> .....	45
Lampiran 7. Kartu Bimbingan Skripsi .....	46
Lampiran 8. Riwayat Hidup .....	48

## DAFTAR PUSTAKA

- Adash, R., Patil, A., Rayar, S., & Veena, K. M. (2019). Comparison of VADER and LSTM for Sentiment Analysis. *International Journal of Recent Technology and Engineering (IJRTE)*, 7(6S), 540–543.
- Anreaja, L. J., Harefa, N. N., Negara, J. G. P., Pribyantara, V. N. H., & Prasetyo, A. B. (2022). Naive Bayes and Support Vector Machine Algorithm for Sentiment Analysis Opensea Mobile Application Users in Indonesia. *JISA(Jurnal Informatika Dan Sains)*, 5(1), 62–68. <https://doi.org/10.31326/jisa.v5i1.1267>
- Bonta, V., Kumaresh, N., & Janardhan, N. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. *Asian Journal of Computer Science and Technology*, 8(S2), 1–6. <https://doi.org/10.51983/ajcst-2019.8.S2.2037>
- Global Media Insight. (2022). *Youtube User Statistic 2022*. [Online] Diakses dari <https://www.globalmediainsight.com/blog/youtube-users-statistics/>.
- Hirzani, F. A., Maharani, W., & Bijaksana, M. A. (2015). Analisis Sentimen Review Produk Menggunakan Pendekatan Berbasis Kamus. *e-Proceeding of Engineering*, 2(2), 5891–5898.
- Ibrohim, M. O., Sazany, E., & Budi, I. (2019). Identify Abusive and Offensive Language in Indonesian Twitter using Deep Learning Approach. *Journal of Physics: Conference Series*, 1196(1), 1–6. <https://doi.org/10.1088/1742-6596/1196/1/012041>
- Isnain, A. R., Sulistiani, H., Hurohman, B. M., & Nurkholis, A. (2022). Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 8(2), 299–303.
- Josi, A., & Abdillah, L. A. (2014). Penerapan Teknik Web Scraping pada Mesin Pencari Artikel Ilmiah. *Jurnal Sistem Informasi (SISFO)*, 5, 159–164.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Chicago: Morgan & Claypool Publishers.
- Manning, C., Raghavan, P., & Schuetze, H. (2009). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

- Martua, E. (2020). *Twitter Covid-19 Sentiment Analysis Lexicon Based*. [Online] Diakses dari <https://github.com/evanmartua34/Twitter-COVID19-Indonesia-Sentiment-Analysis---Lexicon-Based>.
- Muthia, D. A. (2017). Analisis Sentimen pada Review Restoran dengan Teks Bahasa Indonesia Menggunakan Algoritma Naïve Bayes. *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, 2(2), 39–45.
- Nugroho, E. (2011). *Analisa Sentimen Menggunakan Lexicon Based untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Rokok pada Media Sosial Twitter* [Skripsi]. Universitas Brawijaya.
- Nugroho, P. A., Fenriana, I., Ariyanto, R., & Kom, M. (2020). *Implementasi Deep Learning Menggunakan Convolutional Neural Network (CNN) Pada Ekspresi Manusia*. 2(1), 12–21.
- Nurrohmat, M. A., & Sn, A. (2019). Sentiment Analysis of Novel Review Using Long Short-Term Memory Method. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(3), 209–218. <https://doi.org/10.22146/ijccs.41236>
- Nurvania, J., & Lhaksamana, K. M. (2021). Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode Long Short-Term Memory (LSTM). *e-Proceeding of Engineering*, 8(4), 4124–4135.
- Pakpahan, R. (2021). Analisa Pengaruh Implementasi Artificial Intelligence Dalam Kehidupan Manusia. *Journal of Information System, Informatics and Computing*, 5(2), 506–513.
- Prasetyo, E. (2012). *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Penerbit ANDI.
- Pustejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning*. California: O'Reilly.
- Rahman, M. Z., Sari, Y. A., & Yudistira, N. (2021). Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(11), 5120–5127.



- Santoso, A., & Ariyanto, G. (2018). Implementasi Deep Learning berbasis Keras untuk Pengenalan Wajah. *Emitor: Jurnal Teknik Elektro*, 18(1), 15–21. <https://doi.org/10.23917/emitor.v18i01.6235>
- Triawati, C. (2009). *Metode Pembobotan Statistical Concept Based untuk Klustering dan Kategorisasi Dokumen Berbahasa Indonesia* [Skripsi]. Institut Teknologi Telkom.
- Witanto, K. S., Sanjaya ER, N. A., Karyawati, A. E., Kadyanan, I. G. A. G. A., Suhartana, I. K. G., & Astuti, L. G. (2022). Implementasi LSTM Pada Analisis Sentimen Review Film Menggunakan Adam Dan RMSprop Optimizer. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 10(4), 351–362. <https://doi.org/10.24843/JLK.2022.v10.i04.p05>