

## Analisis Sentimen Komentar Youtube Tentang Resesi Global 2023 Menggunakan LSTM

Ari Hendrawan <sup>1</sup>, Enny Itje Sela <sup>2\*</sup>

<sup>1,2</sup> Program Studi Informatika, Fakultas Sains & Teknologi, Universitas Teknologi Yogyakarta, Kabupaten Sleman, Provinsi Daerah Istimewa Yogyakarta, Indonesia.

*Email:* arihendra2802@gmail.com <sup>1\*</sup>, ennysela@uty.ac.id <sup>2</sup>

### Histori Artikel:

*Dikirim* 2 November 2023; *Diterima dalam bentuk revisi* 18 November 2023; *Diterima* 1 Desember 2023; *Diterbitkan* 10 Januari 2024. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

### Abstrak

Jumlah pengangguran meningkat sebagai akibat dari penurunan perekonomian yang disebabkan oleh penurunan aktivitas ekonomi dan pemecatan pekerja oleh perusahaan sebagai akibat dari pandemi COVID-19. Dengan demikian, para aktivis ekonomi memperkirakan bahwa resesi global akan terjadi pada tahun 2023, dan YouTube, platform penyebaran video, menjadi tempat diskusi. Salah satu alasan untuk melakukan analisis sentimen dengan menggunakan data dari komentar video adalah meningkatnya jumlah pengguna YouTube. Analisis sentimen dilakukan dengan Long Short-Term Memory (LSTM). Dengan membagi 500 data menjadi data latih dan data uji, LSTM menghasilkan akurasi tertinggi 90% dan 76% dari data uji, masing-masing. Hasil ini diperoleh dari konfigurasi arsitektur LSTM dengan lapisan tebal yang menggunakan aktivasi sigmoid.

**Kata Kunci:** Analisis Sentimen; Youtube; Resesi Global 2023; Long Short-Term Memory (LSTM).

### Abstract

The COVID-19 pandemic that occurred in 2020 caused the economy to decline due to declining economic activity, making companies decide to lay off some workers so that the unemployment rate increased. This makes economic activists predict that there will be a global recession in 2023, Youtube as a video-sharing platform is one of the places to discuss through the comment's column. The increasing number of YouTube users is one of the references for sentiment analysis using data taken from video comments. Long Short-Term Memory (LSTM) is used to perform sentiment analysis, with 500 data divided into training data and test data, resulting in the highest accuracy of 90% training data and 76% test data. This result is obtained from the configuration of the LSTM architecture with dense layers using sigmoid activation and 50 epochs.

**Keyword:** Sentiment Analysis; Youtube; Global Recession 2023; Long Short-Term Memory (LSTM).

## 1. Pendahuluan

Resesi global adalah keadaan di mana seluruh aktivitas ekonomi global sedang menurun, menyebabkan peningkatan pengangguran dan penurunan nilai mata uang [1]. Pandemi COVID-19 yang terjadi pada 2020 membuat perekonomian menjadi menurun hingga -4,9% yang menggambarkan bahwa resesi global sudah di depan mata [2]. Banyak yang memberikan opini pribadinya melalui media sosial mengenai penurunan aktivitas ekonomi yang terjadi, salah satunya adalah seorang *content creator* (pembuat video) platform Youtube yang memprediksi bahwa tahun 2023 akan terjadinya resesi global. Banyak yang menanggapi video tersebut hanyalah konspirasi dan membawakan ketakutan kepada penonton.

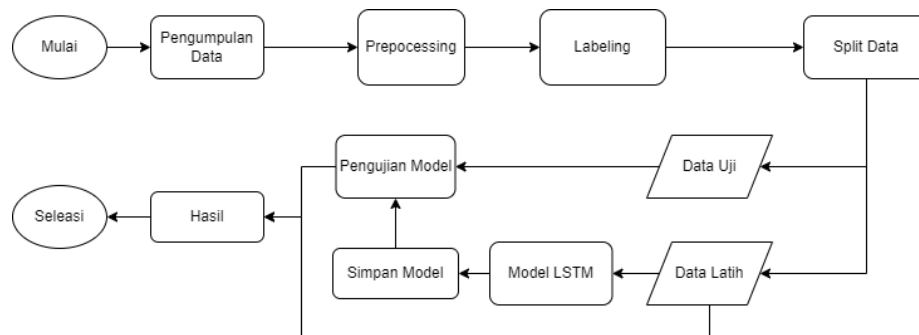
Youtube adalah platform berbagi video yang sangat populer, memungkinkan pengguna menonton dan memuat video [3]. Selain menonton dan memuat Youtube juga dapat melakukan interaksi antara *content creator* dengan penontonya melalui video ataupun kolom komentar, pada kolom komentar penonton dapat memberikan opininya masing-masing mengenai topik videonya. Pengguna Youtube di Indonesia berjumlah 139 juta pengguna [4]. Ini menunjukkan bahwa Youtube dapat menjadi sumber data untuk melakukan analisis sentimen.

Analisis Sentimen adalah analisis pendapat seseorang yang menggambarkan bagaimana merek menyampaikan pendapat seseorang yang menggambarkan bagaimana merek menyampaikan pendapat mereka tentang situasi atau topik yang ditulis. Ini biasanya dibagi menjadi ekspresi senang, tidak senang, dan biasa saja, atau biasanya dibagi menjadi kategori positif, negatif, dan netral [5]. Penelitian terdahulu melakukan analisis sentimen yang dilakukan oleh [6] penelitian tersebut melakukan analisis sentimen menggunakan metode *Naive Bayes Classifier* (NBC) dengan menggunakan data yang diambil menggunakan Twitter API dengan jumlah data sebanyak 1000 data. Terdapat beberapa tahapan yang dilakukan untuk mendapatkan hasil akhir yang diantaranya adalah *preprocessing* yang dilakukan untuk membersihkan data, tahapan *preprocessing* terbagi menjadi beberapa lagi yaitu *cleansing*, *case folding*, *tokenizer*, normalisasi, *filtering*, dan *stemming* kemudian dilakukan pembobotan kata dengan menggunakan TF-IDF. Setelah dilakukan *preprocessing* dan pembobotan kata, dilakukan pelabelan data untuk memberi label pada data dengan dibagi menjadi tiga kelas sentimen yaitu negatif, positif, dan netral. Dengan ini data sudah siap untuk dilakukan pengujian menggunakan metode NBC yang mendapatkan hasil 69%.

Penelitian terdahulu lainnya yang dilakukan oleh [7] melakukan analisis sentimen dengan menggunakan metode *Recurrent Neural Network* (RNN) dan *Naive Bayes*. Dengan menggunakan jumlah 5000 data mendapatkan hasil akurasi tertinggi dengan menggunakan metode RNN dengan pembobotan TF-IDF sebesar 97,77%. ini menunjukkan bahwa metode RNN lebih baik daripada metode *Naive Bayes* yang mendapatkan hasil akurasi 80% dalam penelitian tersebut. RNN memiliki beberapa variasi salah satunya adalah *Long Short-Term Memory* (LSTM) yang memiliki arsitektur lebih kompleks dari pada RNN [8]. Oleh karena itu penulis melakukan penelitian analisis sentimen dengan data yang diambil dari komentar Youtube yang bertopik resesi global 2023 dengan menggunakan metode LSTM.

## 2. Metode Penelitian

Pada penelitian ini akan dilakukan beberapa tahapan untuk melakukan analisis sentimen tanggapan penonton terhadap video mengenai resesi global 2023. Metode *deep learning* akan digunakan pada penelitian ini LSTM yang merupakan salah satu varian dari *Recurrent Neural Networks* (RNN). Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

### 2.1 Pengumpulan Data

Pengumpulan data dilakukan untuk mendapatkan bahan untuk melakukan analisis sentimen. Data yang digunakan adalah data teks yang diambil dari komentar video Youtube yang dengan topik resesi global 2023. Pengumpulan data dilakukan dengan menggunakan YouTube API, yang merupakan kumpulan fungsi yang diberikan oleh Youtube kepada pengembang, yang memungkinkan konten video diintegrasikan ke berbagai perangkat [9]. Sehingga proses pengambilan data komentar Youtube dapat dilakukan melalui Youtube API dengan menggunakan code Python.

### 2.2 Preprocessing

Tahapan *preprocessing* dilakukan untuk membersihkan data yang akan membuat sistem lebih mudah dalam memproses data [10], yang dilakukan dalam tahapan preprocessing adalah *case folding* yang bertujuan untuk mengubah semua teks menjadi huruf kecil, *cleaning* dilakukan untuk menghapus karakter yang tidak diperlukan seperti tanda baca dan emoji. Normalisasi dilakukan untuk mengganti kata-kata singkatan, Tokenisasi dilakukan untuk memecah kalimat menjadi kata, dan *stopword removal* untuk menghapus kata-kata yang tidak diperlukan.

### 2.3 Labeling

Data yang telah dilakukan *preprocessing*, selanjutnya digunakan, pelabelan dengan menggunakan *Textblob*. *Textblob* adalah *library Python* yang digunakan untuk melakukan proses data teks [11], [12]. *TextBlob* akan melakukan pelabelan dengan menghirung polaritas. Selain menggunakan *TextBlob*, dilakukan pelabelan dengan secara manual untuk memastikan data yang diberi label oleh *TextBlob* sudah seimbang.

### 2.4 Split Data

Split data dilakukan untuk membagi data yang sudah diberi label menjadi data latih dan data uji, ini dilakukan dengan secara acak berdasarkan rasio pembagian [13]. Data latih bertujuan untuk melakukan pelatihan pada model LSTM dan data uji bertujuan untuk melakukan pengujian pada model LSTM yang sudah dilatih.

### 2.5 Long Short-Term Memory (LSTM)

*Long Short-Term Memory* (LSTM) adalah variasi dari *Recurrent Neural Network* (RNN). LSTM terdiri dari *cell* yang menerima masukan dan menyimpan nilai selama beberapa waktu, *input gate* mengontrol sejauh mana nilai dari akan masuk ke dalam *cell*, *forget gate* mengontrol seberapa banyak nilai yang ada di dalam *cell*, dan *gate output* mengontrol nilai dalam *cell* yang digunakan untuk menghitung aktivasi keluaran dari unit LSTM [14]. Pada LSTM terdapat beberapa konfigurasi yang diantaranya *epoch*, *layer*, fungsi aktivasi, dan *optimizer*.

### 3. Hasil dan Pembahasan

Pada bagian hasil akan menjelaskan hasil dari proses dan pengujian yang dilakukan. Data yang digunakan pada penelitian ini adalah data teks yang diambil dari komentar Youtube mengenai resesi global 2023 dengan jumlah 500 data. Setelah data didapatkan dilanjutkan pada tahapan preprocessing yang mencakup proses *case folding*, *cleaning*, normalisasi, tokenisasi, dan *stopword removal*. Setelah dilakukan preprocessing jumlah data berkurang menjadi 492, ini dikarenakan terdapat baris yang kosong setelah dilakukan preprocessing. contoh hasil preprocessing dapat dilihat pada Tabel 1.

Table 1. Hasil *Preprocessing*

No	Teks	Teks_bersih
1	Petani yg bakalan hidup tentram	petani yang bakana hidup tentram
2	channel ini kerjanya cuma nakut2in lalu menganjurkan investasi di linknya dia ternyata dagang kecap	channel ini kerjanya cuma nakut in lalu menganjurkan investasi di linknya dia ternyata dagang kecap
3	Jgn percayalah gini ginian. Adanya ini daya beli masyarakat turun. Membuat ekonomi Indonesia melemah	jangan percayalah gini ginian adanya ini daya beli masyarakat turun membuat ekonomi indonesia melemah
4	*Bahkan gado-gado adalah makanan terkuat yang bisa bikin betah seharian* . . . . . _maap, indomi 3ribu adalah raja_	bahkan gado gado adalah makanan terkuat yang bisa bikin betah seharian maap indomi ribu adalah raja

Data yang sudah dibersihkan dengan proses *preprocessing* akan diberi label dengan menggunakan *library python* yaitu *TextBlob* dengan membagi menjadi dua kategori sentimen yaitu positif dan negatif. Selain menggunakan *TextBlob* proses pelabelan juga dilakukan secara manual untuk memastikan bahwa data yang sudah diberi label sudah seimbang. Pada penelitian ini didapati bahwa hasil dari pelabelan menggunakan *TextBlob* mendapati hasil 448 positif dan 42 negatif yang dimana data tersebut tidak seimbang. Untuk memastikan seimbang dilakukan pengubahan label secara manual untuk menyeimbangkan data, hasil dari pelabelan manual adalah 377 positif dan 115 negatif.

Table 2. Hasil *Labeling*

Label	Jumlah
Positif	377
Negatif	115

Selanjutnya dilakukan proses split data untuk membagi data menjadi data latih dan data uji. Data latih adalah data yang digunakan untuk melatih model LSTM yang akan dibuat dan data uji digunakan untuk menguji akurasi dari hasil pelatihan model LSTM. Dari 492 data yang digunakan, data dibagi menjadi 70% data latih dan 30% data uji, yang berarti 344 data latih dan 148 data uji. Selanjutnya dilakukan proses split data untuk membagi data menjadi data latih dan data uji. Data latih adalah data yang digunakan untuk melatih model LSTM yang akan dibuat dan data uji digunakan untuk menguji akurasi dari hasil pelatihan model LSTM. Dari 492 data yang digunakan, data dibagi menjadi 70% data latih dan 30% data uji.

Hasil pengujian dan pelatihan pada analisis sentimen menggunakan model LSTM akan diukur dengan menggunakan nilai akurasi, presisi, *recall*, dan *f1-score*. Untuk melatih dan menguji data terdapat beberapa konfigurasi yang diubah untuk mendapatkan hasil terbaik dari model LSTM. Arsitektur dari model LSTM pada penelitian ini menggunakan *Embedding layer* yang bekerja untuk mengkonversi kata dari data teks menjadi vektor numerik. LSTM *layer* dengan jumlah satu layer dengan 16 unit, *Dropout layer* digunakan untuk mencegah *overfitting* dengan menghilangkan secara acak sebagian unit selama pelatihan dengan mengatur *dropout rate* sebesar 0.5, dan yang terakhir adalah *Dense Layer* dengan 2 unit *output* yaitu 0 atau 1 (negatif atau positif) dengan menggunakan fungsi aktivasi sigmoid dan ReLu yang

dimana kedua aktivasi ini akan diuji sehingga mendapatkan hasil yang terbaik. Selain fungsi aktivasi jumlah *epoch* yang digunakan dalam pelatihan ini adalah 25, 50, dan 100. Hasil dari pelatihan model LSTM dapat dilihat pada Tabel 3.

Table 3. Hasil Pelatihan

Pelatihan	Fungsi Aktivasi	Epoch	Train Acc	Train Loss	Val Acc	Val Loss
1	Sigmoid	25	0.73	0.59	0.78	0.53
2	Sigmoid	50	0.90	0.31	0.74	0.66
3	Sigmoid	100	0.92	0.25	0.71	0.88
4	ReLu	25	0.74	0.91	0.68	0.86
5	ReLu	50	0.78	0.58	0.66	0.63
6	ReLu	100	0.84	0.46	0.71	1.27

Pada hasil pelatihan dapat dilihat bahwa akurasi tertinggi pada data latih adalah 92% dengan menggunakan fungsi aktivasi sigmoid, dan 100 epoch. Untuk membuktikan bahwa hasil dari pelatihan 3 maka dilakukan pengujian dengan data uji, pada hasil pengujian ini akan menampilkan *precision*, *recall*, *f1-score*, *support* dan akurasi. Hasil pengujian dapat dilihat pada Tabel 4.

Table 4. Hasil Pengujian

Pengujian 1				
Kelas	Precision	Recall	F1-Score	Support
Negatif	0.39	0.64	0.49	28
Positif	0.90	0.77	0.83	120
Akurasi	0.74			
Pengujian 2				
Kelas	Precision	Recall	F1-Score	Support
Negatif	0.35	0.73	0.47	22
Positif	0.94	0.76	0.84	126
Akurasi	0.76			
Pengujian 3				
Kelas	Precision	Recall	F1-Score	Support
Negatif	0.37	0.57	0.45	30
Positif	0.87	0.75	0.81	118
Akurasi	0.72			
Pengujian 4				
Kelas	Precision	Recall	F1-Score	Support
Negatif	0.00	0.00	0.00	3
Positif	0.97	0.68	0.80	145
Akurasi	0.67			
Pengujian 5				
Kelas	Precision	Recall	F1-Score	Support
Negatif	0.07	0.30	0.11	11
Positif	0.93	0.69	0.79	138
Akurasi	0.66			
Pengujian 6				
Kelas	Precision	Recall	F1-Score	Support
Negatif	0.43	0.62	0.51	32
Positif	0.88	0.78	0.83	116
Akurasi	0.74			

Berdasarkan pada hasil pengujian yang dilakukan, hasil dari pengujian 2 yang terbaik dengan akurasi 76%. Hasil ini lebih stabil dari pada hasil pengujian 3 yang mendapatkan hasil akurasi pelatihan yang tertinggi dengan hasil 92%. Pada hasil pelatih 2 dengan konfigurasi aktivasi sigmoid dan 50 epoch menghasilkan akurasi tertinggi kedua dengan hasil 90%. Selain itu hasil *f1-score* pada pengujian 2 menaapkatkan hasil tertinggi dengan hasil 84%.

#### 4. Kesimpulan

Berdasarkan hasil penelitian Analisis Sentimen Komentar Youtube Tentang Resesi Global 2023 Menggunakan LSTM, dapat disimpulkan bahwa metode LSTM mendapatkan hasil yang cukup baik untuk melakukan analisis sentiment. Dengan hasil akurasi data latih 90% dan data uji 76%, hasil ini dilakukan dengan menggunakan fungsi aktivasi sigmoid pada dense layer dan melakukan 50 epoch. Nilai akurasi data ujian dan data latih pada penelitian terlihat cukup jauh persentasenya yang bisa diartikan bahwa hasilnya belum konsisten walaupun sudah cukup stabil. Ini bisa diakibatkan karna data yang tidak seimbang dan kurangnya data pada proses pelatihan. Dengan begitu diharapkan penelitian selanjutnya menambahkan data set atau melakukan *oversampling* dan melakukan pengecekan lebih mendalam pada saat preprocessing dan pelabelan.

#### 5. Daftar Pustaka

- [1] Ismail, M., & Indrawati, Y. (2020). *Paradigma Baru Kebijakan Moneter: Menakar Pelajaran Krisis Keuangan Global*. Pustaka Abadi.
- [2] Firdaus, A. H., Listiyanto, E., Talattov, A. P., & Taufikurahman, M. R. (2020). *Kajian tentang tahun INDEF 2020: Menata arsitektur ekonomi pasca pandemi*. INDEF.
- [3] Sari, L. (2020). Upaya menaikkan kualitas pendidikan dengan pemanfaatan youtube sebagai media ajar pada masa pandemi covid-19. *Jurnal Tawadhu*, 4(1), 1074-1084.
- [4] Sadya, S., & Bayu, D. (2023). Daftar Negara Pengguna YouTube Terbesar Awal 2023, Ada Indonesia. *DataIndonesia. id. Feb*, 24, 2023-11.
- [5] Nugraha, F. A., Harani, N. H., & Habibi, R. (2020). *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Kreatif.
- [6] Novitasari, I., Kurniawan, T. B., & Dewi, D. A. (2022). Analisis Sentimen Masyarakat Terhadap Tweet Ruang Guru Menggunakan Algoritma Naive Bayes Classifier (NBC). *Jurnal Mantik*, 6(3), 3308-3318.
- [7] Lestandy, M., Abdurrahim, A., & Syafa'ah, L. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naive Bayes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 802-808. DOI: <https://doi.org/10.29207/resti.v5i4.3308>.
- [8] Kumar, A., Sangwan, S. R., Arora, A., Nayyar, A., & Abdel-Basset, M. (2019). Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network. *IEEE access*, 7, 23319-23328. DOI: 10.1109/ACCESS.2019.2899260.

- [9] Fatmasari, R., Ayu, V. M., Anto, H., & Yulianto, L. D. (2022). Analisis Sentimen Dalam Pengkategorian Komentar Youtube Terhadap Layanan Akademik dan Non-Akademik Universitas Terbuka Untuk Prediksi Kepuasan. *Building of Informatics, Technology and Science (BITS)*, 4(2), 395-404. DOI: <https://doi.org/10.47065/bits.v4i2.1738>.
- [10] Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan metode naïve bayes dan support vector machine pada analisis sentimen twitter. *SMATIKA Jurnal: STIKI Informatika Jurnal*, 10(02), 71-76. DOI : <https://doi.org/10.32664/smatika.v10i02.455>.
- [11] Azhar, R., Surahman, A., & Juliane, C. (2022). Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 6(1), 267-281. DOI: <http://dx.doi.org/10.30645/j-sakti.v6i1.443>.
- [12] Parlita, R., Pradika, S. I., Hakim, A. M., & NM, K. R. (2020). Analisis Sentimen Twitter Terhadap Bitcoin dan Cryptocurrency Berbasis Python TextBlob. *Jurnal Ilmiah Teknologi Informasi dan Robotika*, 2(2), 33-37.
- [13] Setiawati, I., & Sela, E. I. (2022). Classification of Facial Expression Using Principal Component Analysis (PCA) Method and Support Vector Machine (SVM). *International Journal of Computer and Information Technology (2279-0764)*, 11(1).
- [14] Khumaidi, A., Raafi'udin, R., & Solihin, I. P. (2020). Pengujian Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung. *Jurnal Telematika*, 15(1), 13-18. DOI: <https://doi.org/10.61769/telematika.v15i1.340>.