

Analisis Sentimen Pengguna Youtube Mengenai Analog Switch Off Menggunakan Word Embedding Dan Metode Long Short-Term

Mochamad Suhri Ainur Rifky¹, Amalia Anjani Arifiyanti²,
Reisa Permatasari³

^{1,2,3} Universitas Pembangunan Nasional "Veteran" Jawa Timur

Alamat: Jl. Rungkut Madya No.1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294

Korespondensi penulis: mochamad.suhri@gmail.com

Abstract. Analog Switch Off (ASO) or migration program from analog television to digital television is a program issued by the Ministry of Communication and Informatics in Indonesia. Some people provide different responses and opinions on YouTube comments about ASO. There are those who give positive or neutral comments. However, there were also those who gave negative comments. Sentiment analysis is a process that is carried out automatically in studying, retrieving, and processing textual data to obtain information and see responses or opinions about an issue or object towards positive, neutral or negative opinions. Thus sentiment analysis can be used as a reference in making organizational decisions, improving a service, or as a review of a product. Sentiment analysis was performed using word embedding with Word2Vec, and sentiment classification using the Long Short-Term Memory (LSTM) method. The results of the evaluation test are 92% accuracy, 92% precision, 92% recall, and 92% f1-score.

Keywords: Long Short-Term Memory, Sentiment Analysis, Word Embedding,

Abstrak. Analog Switch Off (ASO) atau program migrasi dari televisi analog ke televisi digital merupakan program yang dikeluarkan oleh Kementerian Komunikasi dan Informatika di Indonesia. Beberapa masyarakat memberikan tanggapan dan opini yang berbeda-beda di komentar YouTube mengenai ASO. Ada yang memberikan komentar bersetimen positif atau netral. Namun, ada juga yang memberikan komentar bersetimen negatif. Analisis sentimen adalah proses yang dilakukan secara otomatis dalam mempelajari, mengambil, dan mengolah data tekstual untuk memperoleh informasi dan melihat tanggapan atau opini mengenai sebuah masalah atau objek menuju opini positif, netral atau negatif. Maka sebab itu analisis sentimen dapat digunakan sebagai acuan dalam pengambilan kebijakan organisasi, peningkatan suatu layanan, atau sebagai bahan evaluasi produk. Analisis sentimen dilakukan menggunakan *word embedding* dengan Word2Vec, serta pengklasifikasian sentimen menggunakan metode *Long Short-Term Memory* (LSTM). Uji evaluasi menghasilkan *accuracy* 92%, *precision* 92%, *recall* 92%, dan *f1-score* 92%.

Kata kunci: Long Short-Term Memory, Analisis Sentimen, Word Embedding

LATAR BELAKANG

Salah satu media informasi yang penting yaitu televisi, dikarenakan bisa menyampaikan informasi dengan cepat dan efektif. Kementerian Komunikasi dan Informatika (Kemkominfo) bersama Katadata Insight Center (KIC) melakukan survei mengenai "Ragam Sumber Informasi yang Paling Banyak Diakses Masyarakat" menunjukkan bahwa, setelah media sosial, televisi menempati peringkat kedua sebagai sumber informasi utama bagi masyarakat Indonesia dengan minat yang paling tinggi (Annur, 2022). Di Indonesia digunakan dua jenis sistem siaran televisi yaitu sistem televisi analog dan digital. Namun, dengan berkembangnya teknologi, penggunaan televisi analog dianggap kurang modern (Andarini, 2022). Oleh karena itu,

Kemkominfo telah mengeluarkan program untuk mengakhiri sistem televisi analog dan melakukan migrasi ke televisi digital.

Proses migrasi televisi analog menjadi televisi digital merupakan salah satu perkembangan penting dan solusi alternatif di Indonesia saat ini, terutama setelah diterbitkannya Undang-Undang Nomor 11 Tahun 2020 tentang Cipta Lapangan Kerja yang disahkan oleh Presiden Indonesia Joko Widodo pada tanggal 2 November 2020. Hal ini menunjukkan komitmen pemerintah dalam sektor industri penyiaran di Indonesia. Undang-Undang tersebut menegaskan posisi Indonesia dalam menyambut era penyiaran digital. Ada perintah untuk segera melakukan digitalisasi penyiaran yang tertuang dalam Undang-Undang Nomor 11 Tahun 2020 tentang Cipta Lapangan Kerja pada Pasal 60A. Pasal tersebut menjadi dasar hukum untuk berlakunya migrasi penyiaran analog ke penyiaran digital atau disebut juga sebagai *Analog Switch Off* (ASO) (Ramadhan, 2021).

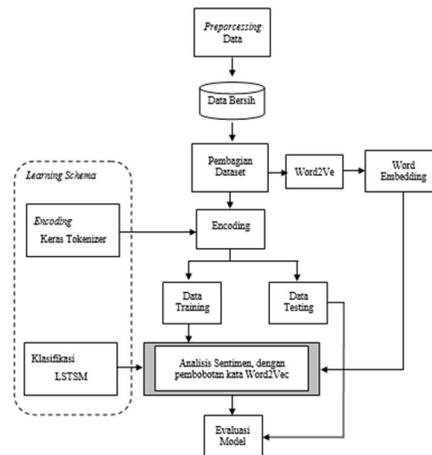
Terhitung sampai awal tahun 2023, Proses pemberhentian siaran televisi analog atau ASO sudah dilakukan di 265 wilayah dari total 514 wilayah (Wisnubroto, 2023). Beberapa masyarakat memberikan tanggapan dan opini yang berbeda-beda di media sosial mengenai ASO. Ada yang mengungkapkan pendapat positif di komentar Youtube dan merasa puas setelah melakukan migrasi. Namun, ada juga yang memberikan tanggapan dan opini negatif, mengeluhkan hilangnya siaran dan jumlah siaran yang sedikit pada TV digital melalui komentar Youtube.

Pendapat dan komentar dapat mencerminkan pandangan positif atau negatif melalui analisis sentimen. Oleh karena itu, hasil analisis sentimen dapat berperan sebagai acuan dalam pengambilan kebijakan organisasi, peningkatan suatu layanan, atau sebagai bahan evaluasi produk. Analisis sentimen yaitu proses yang dilakukan secara otomatis dalam mempelajari, mengambil, dan mengolah data tekstual untuk memperoleh informasi dan melihat tanggapan atau opini mengenai sebuah masalah atau objek menuju opini positif atau negatif (Rahutomo, 2018).

Salah satu bagian metode dari *Deep Learning* yaitu *Long Short-Term Memory* (LSTM) mampu dimanfaatkan untuk *Natural Language Processing* (NLP), seperti analisis sentimen, pengenalan suara, dan translasi teks. LSTM adalah pengembangan dari metode *Recurrent Neural Network* (RNN) yang diciptakan untuk mengatasi masalah yang ditemukan pada RNN, yaitu vanishing gradient. Penelitian (Hassan & Mahmood, 2017) dan (Wang & Liu, 2018) menunjukkan bahwa metode LSTM memiliki tingkat akurasi yang lebih tinggi daripada metode konvensional. Dalam melakukan proses yang berkaitan dengan NLP, ada baiknya jika

mampu mengetahui hubungan antar kata di dalamnya. Penerapan *word embedding* dapat merepresentasikan mengenai hubungan antar kata dalam bentuk ruang vektor kontinu (Wang & Liu, 2018).

METODE PENELITIAN



Gambar 1 Metode Penelitian

Preprocessing

Proses awal dalam *text mining* adalah *preprocessing*, yang bertujuan untuk menyaring data yang akan diproses sehingga menghasilkan data yang lebih optimal dan terstruktur dengan jelas (Balya, 2019). Tahapan dalam *Preprocessing* meliputi pembersihan data (*cleaning*), mengubah seluruh teks menjadi huruf kecil (*case folding*), menghapus kata-kata pengisi yang tidak penting (*stopword removal*), dan mengubah kata-kata ke kata dasarnya (*stemming*) (Maisarah, 2020).

Pembagian Dataset

Pembagian komposisi dataset untuk training dan testing model LSTM yang dibuat menggunakan rekomendasi dengan komposisi sebesar 80%:20%. Dengan catatan, porsi dataset untuk proses training yang lebih besar.

Word Embedding dengan Word2vec

Word embedding adalah suatu metode untuk merepresentasikan kata-kata dengan makna serupa dalam bentuk representasi yang serupa pula (Brownlee, 2017). Pada tahap ini bertujuan untuk merubah setiap kata pada dataset training menjadi nilai vector agar dapat diproses dalam model LSTM.

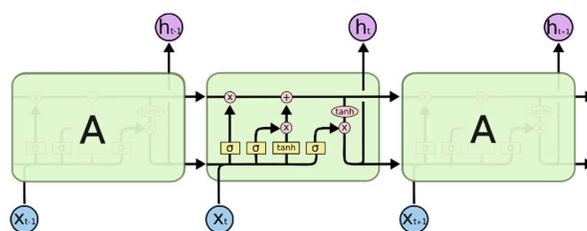
Encoding menggunakan Keras Tokenizer

Encoding merupakan teknik untuk menggambarkan sebuah kata sebagai vektor numerik. Pada langkah ini, dilakukan proses tokenize pada data untuk mengubahnya menjadi format

yang dapat dimanfaatkan oleh word embedding. Dalam Keras, disediakan berbagai metode kenyamanan untuk melakukan preprocessing pada teks dan urutan data, sehingga siap digunakan. Salah satu metode yang bisa digunakan adalah menggunakan kelas utilitas Tokenizer, yang berfungsi untuk mengonversi kumpulan teks menjadi daftar bilangan integer. Setiap bilangan integer ini kemudian dipetakan ke nilai tertentu dalam kamus yang mengkodekan seluruh korpus teks.

Klasifikasi Sentimen Model LSTM

Setelah dilakukan *word embedding* dan *encoding*, selanjutnya dilakukan adalah training model LSTM dengan menggunakan data yang telah dilabeli pada tahapan sebelumnya. Training model ini bertujuan untuk mengoptimalkan performa model agar mampu mengelompokkan data dengan tepat sesuai dengan karakteristik.



Gambar 2 LSTM

Dalam arsitektur LSTM, terdapat *cell state* yang memiliki kemampuan untuk menyimpan informasi dalam jangka waktu lama atau singkat pada sel LSTM. Meskipun secara teoritis RNN dapat menangani ketergantungan jangka panjang antar data, namun RNN tidak dapat secara otomatis memilih parameter yang tepat untuk menyelesaikan masalah tertentu. Hal ini menyebabkan RNN tidak dapat belajar atau memahami masalah tersebut dengan baik (Seo et al., 2020).

Fungsi aktivasi sigmoid pertama dalam sel LSTM disebut *forget gate* berfungsi untuk menentukan informasi mana yang harus dihapus dari keadaan sel sebelumnya ($ct - 1$). Sigmoid kedua berfungsi sebagai *input gate* yang mengontrol informasi baru yang masuk ke dalam sel. Sedangkan sigmoid terakhir bertindak sebagai *output gate* yang menentukan informasi mana yang harus diteruskan ke status tersembunyi (*hidden status*).

Evaluasi Model

Evaluasi Model merupakan proses analisis sentimen yang dilakukan dalam penelitian ini. Pada tahap ini, hasil dari evaluasi akan ditampilkan dalam bentuk nilai tertentu yang dapat merepresentasikan performa model terhadap atribut yang telah dipersiapkan sebelumnya.

HASIL DAN PEMBAHASAN

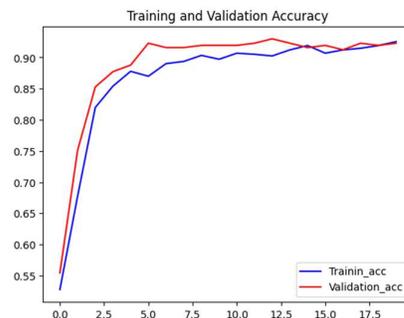
Bagian ini memuat hasil dari penelitian yang telah dilakukan berdasarkan dengan metode yang telah diuraikan. Dalam melakukan penelitian ini digunakan bahasa pemrograman python. Kemudian dataset yang digunakan adalah data yang diambil dari komentar YouTube dengan topik ASO, yang dikumpulkan dengan cara *crawling*. Selanjutnya dilakukan evaluasi menggunakan *confusion matrix* dan *classification report*.

Dataset yang telah dikumpulkan dan telah dilabeli akan melalui tahap *preprocessing*, yaitu yaitu *cleaning*, *case folding*, *stopword removal* dan *stemming*. Sehingga selanjutnya menghasilkan data bersih yang akan digunakan untuk pelatihan model dan evaluasi model. Dataset dibagi menjadi dua dengan komposisi 80% untuk data pelatihan dan 20% untuk data evaluasi.

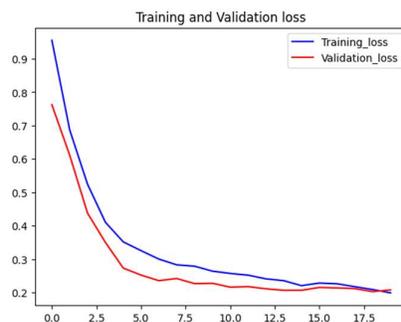
Tabel 1 *Preprocessing*

Data sebelum <i>preprocessing</i>	Data sesudah <i>preprocessing</i>
TV DIGITAL ... JUSTRU BIKIN STREEES ... MASALAH SINYAL PARAAH	digital strees sinyal paraah
Alhamdulillah saya sudah pake tv digital dgn bantuan alat set Top box..dapet 51 Chanel.. gambarnya jernih dengan antena digital out door	alhamdulillah pake digital bantu alat set top box dapet chanel gambar jernih antena digital out door
Siaran buruk, mulai dari era Covid, sudah menontonnya, lebih baik menonton YouTube	siar buruk era covid tonton tonton youtube

Tahap klasifikasi sentimen menggunakan model LSTM menggunakan data sebanyak 1809 yang telah dilakukan *word embedding* menggunakan Word2Vec dan juga telah di-*encoding*. Pada tahap ini dilakukan pelatihan sebanyak 20 *epoch* pada model LSTM, hasil pelatihan model LSTM ditunjukkan dalam grafik *Training and Validation Accuracy* dan *Training and Validation Loss*



Gambar 3 Grafik Training and Validation Accuracy



Gambar 3 Training and Validation Loss

Setelah dilakukan pelatihan model sebanyak 20 *epoch*, didapatkan bahwa performa dari model mampu melakukan pembelajaran dengan baik menggunakan dataset *training*. Model berhasil melakukan klasifikasi dengan *training accuracy* and *validation accuracy* sebesar 0.9252 dan 0.9288 pada *epoch* 20, sedangkan *training loss* and *loss accuracy* sebesar 0.1989 dan 0.2077 pada *epoch* 20.

Dengan melakukan perbandingan nilai *accuracy* dan *loss* pada proses pelatihan, model belum menghasilkan hasil klasifikasi sebenarnya. Oleh karena itu, perlu dilakukan evaluasi model klasifikasi. Evaluasi model dilakukan menggunakan dataset *testing* untuk mengevaluasi apakah model klasifikasi menghasilkan *accuracy* yang baik atau tidak. Cara melakukan evaluasi model ini yaitu dengan menggunakan *confusion matrix* dan *classification report*.

Tabel 2 Confusion Matrix

	Prediksi Negatif	Prediksi Netral	Prediksi Positif
Sebenarnya Nergatif	123	7	4
Sebenarnya Netral	6	78	0
Sebenarnya Positif	6	4	128

	precision	recall	f1-score	support
0	0.91	0.92	0.91	134
1	0.88	0.93	0.90	84
2	0.97	0.93	0.95	138
accuracy			0.92	356
macro avg	0.92	0.92	0.92	356
weighted avg	0.93	0.92	0.92	356

Gambar 4 Classification Report

Berdasarkan tabel *confusion matrix* dan gambar *classification report*, model mampu melakukan prediksi dengan baik menggunakan data *testing* dengan menghasilkan *accuracy* 92%, *precision* 92%, *recall* 92%, dan *f1-score* 92%.

KESIMPULAN DAN SARAN

Berdasarkan hasil pelatihan dan evaluasi klasifikasi sentimen menggunakan *word embedding* yaitu Word2Vec dan metode *Long Short-Term Memory* (LSTM) pada dataset yang diambil dari komentar pengguna YouTube berbahasa Indonesia dengan topik *Analog Switch Off* (ASO) menghasilkan performa yang baik. Pada tahap pelatihan, model menghasilkan *training accuracy* and *validation accuracy* sebesar 0.9252 dan 0.9288 pada *epoch* 20, serta *training loss* and *loss accuracy* sebesar 0.1989 dan 0.2077 pada *epoch* 20. Sedangkan pada tahap evaluasi, model menghasilkan *accuracy* 92%, *precision* 92%, *recall* 92%, dan *f1-score* 92%.

Dalam pengembangan kedepannya dapat dilakukan penambahan tahapan dalam proses *preprocessing* yaitu *slang word*, serta dapat dilakukan penambahan *list stopwords* ketika melakukan *stopword removal*.

DAFTAR REFERENSI

- Annur CM. (2022) Survei KIC: Mayoritas Masyarakat Indonesia Mengakses Informasi di Media Sosial. Diakses pada 15 Januari 2023, dari <https://databoks.katadata.co.id/datapublish/2022/01/20/survei-kic-mayoritas-masyarakat-indonesia-mengakses-informasi-di-media-sosial>
- Balya. (2019). Analisis Sentimen Pengguna Youtube di Indonesia pada Review Smartphone Menggunakan Naïve Bayes. Medan: USU.
- Brownlee, J. (2017). Deep Learning for Natural Language Processing. Diakses pada 23 Januari 2023, https://github.com/abhiramkadali/Machine_Learning-_Resources.
- Hassan, A., & Mahmood, A. (2017). Deep learning for sentence classification. Farmingdale. USA.
- Hochreiter, S. & Chmidhuber, J. U. (1997). Long Short-Term Memory, Neural Computation. *Neural Computation*, 9(8), 1735-1780.
- Maisarah, M. A. (2020). Sistem Analisis Sentimen pada Fanpage Facebook Kandidat Presiden 2019-2024.
- Rahutomo, F., Saputra, P. Y., & Fidyawan M. A. (2018) Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine. *Jurnal Inform Polinema*, 4(2), 93–100.
- Ramadhan, M. H. (2021), *Analog Switch Off Dalam Perspektif Regulator (Studi Kasus Peranan Komisi Penyiaran Indonesia Pusat Dalam Pemberlakuan Analog Switch Off Pasca Penetapan Undang-Undang Nomor 11 Tahun 2020)*. Bandung: Universitas Komputer Indonesia.

- Seo, S., Kim, C., Kim, H., Mo, K., & Kang, P. (2020). Comparative Study of Deep Learning-Based Sentiment Classification. *IEEE Access*, 8, 6861–6875.
- Su, Z., Xu, H., Zhang, D., & Xu, Y. (2014). Chinese sentiment classification using a neural network tool — Word2vec. *International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI)*, Beijing, China, pp. 1–6.
- Wang, J. H., & Liu, T. W. (2018). An LSTM Approach to Short Text Sentiment Classification with Word Embeddings. *The 2018 Conference on Computational Linguistics and Speech Processing*, pp. 214-223.
- Wisnubroto, K. (2023). Agar Keinginan Nonton Siaran TV Digital tak Tersandung Harga STB. Diakses pada 20 Januari 2023, dari <https://www.indonesia.go.id/kategori/editorial/6808/agar-keinginan-nonton-siaran-tv-digital-tak-tersandung-harga-stb?lang=1>.