



ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA LSTM PADA MEDIA SOSIAL

Andi Aljabar^a, Arifin A Abd Karim^b

^a Fakultas Teknik dan Ilmu Komputer / Teknik Informatika, aljabar.android@unusia.ac.id, Unusia

Fakultas Teknik dan Ilmu Komputer / Teknik Informatika, arifinkarim@unusia.ac.id, Unusia

ABSTRAK

Social media produces a lot of information such as information related to politics, social, culture, sports and even knowledge. Not only that, many people comment on an event from a certain television station through social media. These comments can be positive or negative comments. On the other hand, television stations also need this to be the basis for designing an event that is liked and has positive value by the community. One of the machine learning methods that can be used to support this is LSTM (Long Short Term Memory). The main purpose of this research is how to use the LSTM method in determining the form of a commentary in the form of comments on the television program, whether the event is sentimental or vice versa. Based on the test results using the LSTM method, it shows an accuracy rate of 97% and a loss of 12%..

Keywords: LSTM, Machine Learning, Algoritma LSTM.

Abstrak

Sosial media menghasilkan banyak informasi seperti informasi terkait politik, social, budaya, olahraga bahkan pengetahuan. Tidak hanya itu, banyak dari masyarakat yang mengomentari suatu acara dari stasiun televisi tertentu melalui social media. Komentar-komentar tersebut bisa berupa komentar positif atau negative. Disisi lain, stasiun televisi juga juga membutuhkan hal tersebut guna menjadi dasar dalam merancang suatu acara yang disukai dan bernilai positif oleh masyarakat. Salah satu metode machine learning yang dapat digunakan untuk menunjang hal tersebut adalah LSTM (Long Short Term Memory). Tujuan utama pada penelitian ini adalah bagaimana penggunaan metode LSTM dalam menentukan bentuk dari suatu komentar berupa komentar terhadap acara televisi tersebut, apakah acara tersebut bermuatan sentiment atau sebaliknya. Berdasarkan hasil pengujian menggunakan metode LSTM memperlihatkan tingkat akurasi sebesar 97% dan loss sebesar 12%..

Kata Kunci: LSTM, Machine Learning, Algoritma LSTM

1. PENDAHULUAN

Berdasarkan data We Are Social di tahun 2020 terdapat sekitar 175,4 juta user yang menggunakan internet di Indonesia. Angka ini meningkat dibandingkan dengan tahun sebelumnya yaitu 145,2 juta. Total jumlah penduduk Indonesia berkisar 272,1 juta jiwa, dengan demikian pengguna internet di Indonesia mencapai angka sekitar 64% (Tri Hermanto et al., n.d.). Dengan pertumbuhan teknologi yang begitu pesat ditambah banyaknya aplikasi social media sehingga merubah pola komunikasi terhadap sesama masyarakat. Penggunaan media sosial juga banyak dimanfaatkan oleh masyarakat untuk menyampaikan opini dan saling berbagi antara sesama.

Perkembangan teknologi yang begitu pesat terutama dalam hal sosial media juga, secara tidak langsung mengumpulkan banyak data atau big data yang dimana setiap individu atau pengguna internet dapat menyampaikan opininya dan juga sikapnya terhadap suatu hal (Abu Farha & Magdy, 2021). Sebagai contoh terhadap suatu acara atau tayangan di televisi maupun media internet, setiap individu atau yang sering kita sebut netizen dapat menyampaikan opininya melalui sosial media terhadap tayangan tersebut. Opini ini

Received Juni 30, 2022; Revised Juli 2, 2022; Accepted Agustus 22, 2022

berupa data yang dapat dijadikan acuan untuk meningkatkan kualitas dari tayangan tersebut baik dari sisi pendidikan, gambar dan lain sebagainya.

Tayangan televisi sangat berpengaruh terhadap dampak dan perilaku terhadap masyarakat terutama bagi anak-anak (Sabardila et al., 2021). Oleh karena itu dengan pemanfaatan social media, dapat dikumpulkan komentar-komentar masyarakat terhadap tayangan televisi serta dapat ditarik kesimpulan terhadap tayangan televisi tersebut. Disisi lain, pengolahan data tersebut secara manual membutuhkan waktu dan tenaga yang tidak sedikit, sehingga dibutuhkan suatu teknologi yang dapat mengolah dan mempermudah proses pengolahannya.

Sentimen analisis merupakan suatu proses otomatisasi yang digunakan untuk membantu dalam proses pengklasifikasian konten apakah konten tersebut objektif atau opini (Jianqiang et al., 2018). Beberapa penelitian telah melakukan pengkajian terkait sentiment analisis dari beberapa objek yang ada disosial media seperti pengakuan emosi menggunakan metode MLP dan LSTM dengan dengan hasil akurasi yang hampir sama yaitu 71,74%, dimana MLP lebih unggul dalam tahap pemrosesannya (Tri Hermanto et al., n.d.). CNN dan LSTM dengan objek bahasa isyarat Indonesia atau BISINDO 96% (Aljabar & Suharjito, 2020). Sentiment dengan isu politik menggunakan metode Naïve Byesian 92% (Kumar & Jaiswal, 2020).

Dari beberapa penelitian sebelumnya, terdapat perbedaan tingkat akurasi, hal ini disebabkan penggunaan model atau metode learning serta perbedaan objek atau dataset yang digunakan. Sehingga menarik peneliti untuk melakukan penelitian dengan objek sentiment analisis pada acara televisi terutama di Indonesia.

2. TINJAUAN PUSTAKA

Biasanya, riset terkait sentiment analisis tradisional menggunakan machine learning digunakan sebagai core klasifikasi atau kluster data (Basiri et al., 2021). Selain itu informasi sentiment analisis pada media sosial juga menarik perhatian para peneliti karena dapat merangsang pola pikir nitizen untuk memberikan pendapatnya terhadap suatu objek (Wang et al., 2020). Seperti yang telah dibahas sebelumnya, dengan begitu banyaknya data atau komentar-komentar di-internet, ada begitu banyak data yang bisa diolah atau diteliti salah satunya adalah opini masyarakat terhadap suatu acara yang ditayangkan pada stasiun televisi.

2.1. Dataset dan Preprocessing

Dataset merupakan data yang digunakan untuk menentukan suatu model pada machine learning (Aljabar & Suharjito, 2020). Dataset tentang opini masyarakat terkait tayanagan suatu acara distasiun televisi dikumpulkan melalui social media twitter (Nurjanah et al., 2017). Dataset tersebut berfungsi sebagai acuan untuk melakukan trining dan testing data dan selanjutnya akan menjadi sebuah model yang digunakan untuk melakukan prediksi sentiment analisis terhadap suatu kalimat.

Preprocessing merupakan proses pembersihan data atau field-field pada data yang tidak digunakan pada saat proses trining, validation, dan testing (Aljabar & Suharjito, 2020). Proses preprocessing sering juga disebut cleaning juga merupakan proses pembersihan data seperti membersihkan karakter-karakter khusus (emote, special karakter) (Tri Hermanto et al., n.d.). Adakalanya pada dataset terdapat banyak field, sedangkan yang dibutuhkan hanya terdiri dari beberapa field. Sehingga field-field atau data-data yang tidak complete sebaiknya dibersihkan terlebih dahulu karena hal tersebut sangat berpengaruh terhadap proses trining dan testing.

2.2. Pembobotan

Pembobotan bertujuan memberi label positif atau negative value dari setiap kalimat yang ada pada dataset (Wang et al., 2020). Pembobotan ini dilakukan secara manual untuk mengetahui jenis kalimat seperti apa yang dikategorikan sebagai opini sentiment. Metode yang dilakukan untuk melakukan pelabelan adalah majority voting (Tri Hermanto et al., n.d.).

2.3. Tokenize

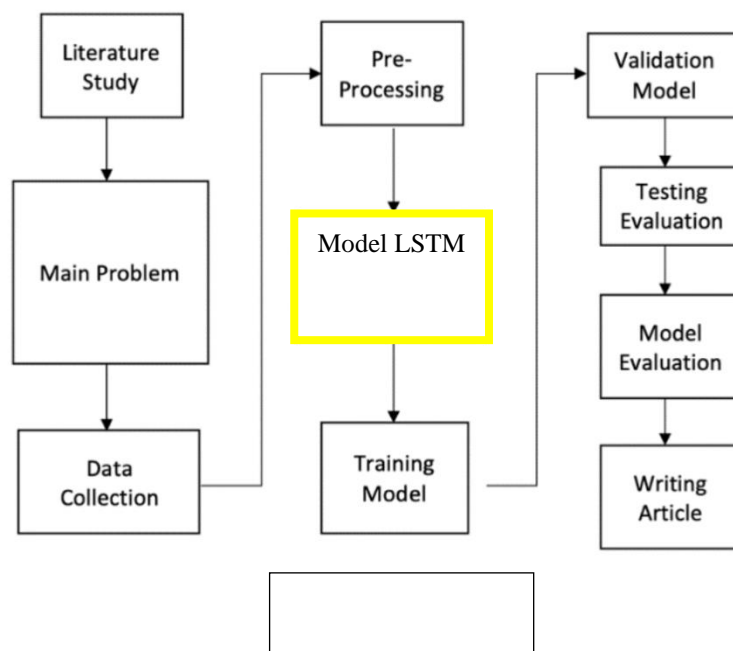
Langkah pertama sebelum trining dan testing adalah tokenize yaitu memberikan token atau nilai khusus pada masing-masing kata dimasing-masing dataset. Biasanya proses tokenize ini memisahkan kata per-kata berdasarkan karakter spasi (Tri Hermanto et al., n.d.). Setelah proses tokenize dilakukan maka kalimat-kalimat tadinya berupa kata per-kata selanjutnya berubah dalam bentuk angka dan tersusun menjadi array yang selanjutnya akan diolah atau dilakukan training dan testing.

2.4. LSTM

LSTM merupakan salah satu metode yang banyak digunakan untuk mengolah kata pada NLP (Neural Language Processing). LSTM banyak digunakan pada data yang bersifat kata atau text base karena sifatnya yang sequential (Aljabar & Suharjito, 2020). Kerangka utama pada LSTM mempunyai tiga gerbang yaitu input, forget, dan output gate (Tri Hermanto et al., n.d.). LSTM juga merupakan model pembaharuan dari RNN (Recurrent Neural Network), dengan pengembangan memory cell (Basiri et al., 2021).

3. METODOLOGI PENELITIAN

Langkah awal adalah melakukan studi literatur mengenai pengaruh acara televisi terhadap kehidupan sosial masyarakat, melalui social media dan juga studi literatur terkait machine learning, neural language processing dan beberapa metode terkait hal tersebut seperti CNN, RNN, K-Nearest, dan LSTM. Selanjutnya adalah melakukan data collection atau pengumpulan data. Setelah data terkumpul maka dilakukan preprocessing dan pelabelan data. Setelah data memiliki label maka proses berikutnya adalah melakukan tokenize, trining, validation dan testing model menggunakan LSTM. Terakhir melakukan evaluasi model dan penulisan artikel. Untuk lebih jelasnya seperti yang tertuang dalam gambar berikut.



Gambar 1. Model LSTM [1]

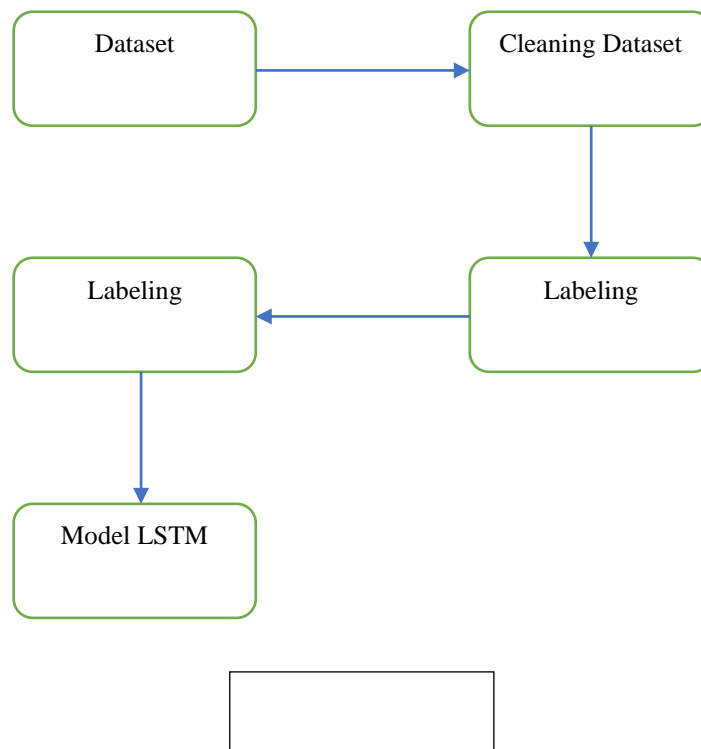
Seperti yang telah dikatakan sebelumnya, ada beberapa referensi terkait metode-metode yang ada pada machine learning, seperti CNN+LSTM dengan object sign language dengan akurasi 96% (Aljabar & Suharjito, 2020). CNN 79.68%, LSTM 80.93% dengan objek sentiment analisis text bahas China (Feng & Cheng, 2021). Prediksi pemilihan presiden berdasarkan pengolahan hasil dari sosial media (Budiharto & Meiliana, 2018). K-Nearest dengan objek text Bahasa Indonesia 82.50% (Nurjanah et al., 2017). Adapun tabel perbandingan metode dan berdasarkan objek sebagai berikut :

Tabel 1. Tabel Perbandingan Metode Berdasarkan Objek

Model	Objek	Akurasi
CNN	Bahasa Isyarat Indonesia	96%
LSTM	Bahasa Isyarat Indonesia	86%
CNN + LSTM	Bahasa Isyarat Indonesia	96%
CNN	Text Bahas China	79,68
LSTM	Text Bahas China	80,93%
K-Nearest	Text Bahasa Indonesia	82,50%

3.1. Metode Usulan

Langkah pertama yang dilakukan adalah melakukan pengumpulan dataset yang terdiri dari 400 row. Setelah dataset terkumpul maka langkah selanjutnya adalah melakukan cleaning data yaitu dengan menghapus kolom yang tidak dibutuhkan serta menghapus karakter-karakter yang tidak dibutuhkan serta memberikan label atau kategori positif dan negative pada data tersebut. Selanjutnya adalah tahap tokenize, dimana setiap kata memiliki nilai value berupa angka. Untuk lebih jelasnya terdapat pada gambar berikut :

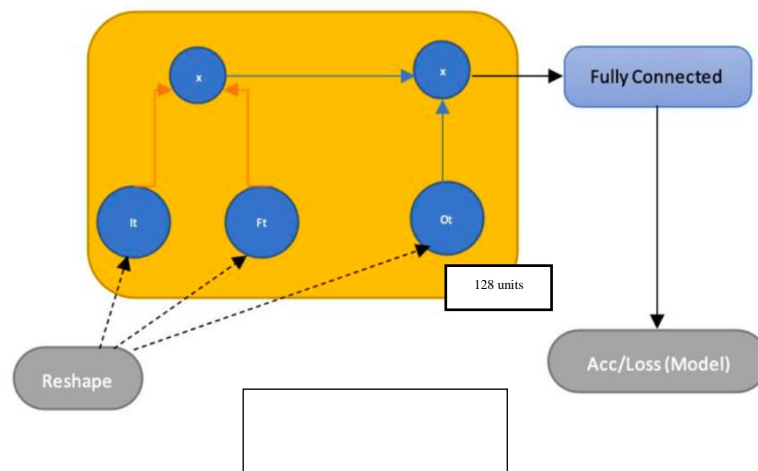


Gambar 2. Proses Tokenize [2]

Pada proses tokenize, apabila terdapat kata yang sama, maka kata tersebut memiliki nilai value yang sama. Sebagai contoh, apabila terdapat kata “tidak” dan diberi nilai value “5” maka setiap kali model menemukan kata tidak pada row yang berbeda dan kata “tidak” pada masing-masing baris tersebut akan bernilai “5”. Selanjutnya value atau angka tersebut akan disusun menjadi sebuah matrix dan selanjutnya akan dilakukan proses validation dan testing. Dengan memanfaatkan atribut “test_size” Dataset yang terdiri dari 400 kata dibagi menjadi 2 kelompok yaitu 75% sebagai data training dan 25% sebagai data validation testing.

3.2. LSTM (Long Short Term Memory)

Gambar berikut menjelaskan bagaimana LSTM bekerja. Proses input dari reshape selanjutnya akan dikirim ke gate, apabila terdapat noise pada data maka data tersebut akan dikirim ke forgot get untuk dibersihkan terlebih dahulu, apabila data tersebut sudah clean maka data akan disimpan ke stored gate. Langkah tersebut dilakukan terus menerus sebanyak 128 kali sesuai dengan units yang digunakan pada LSTM.

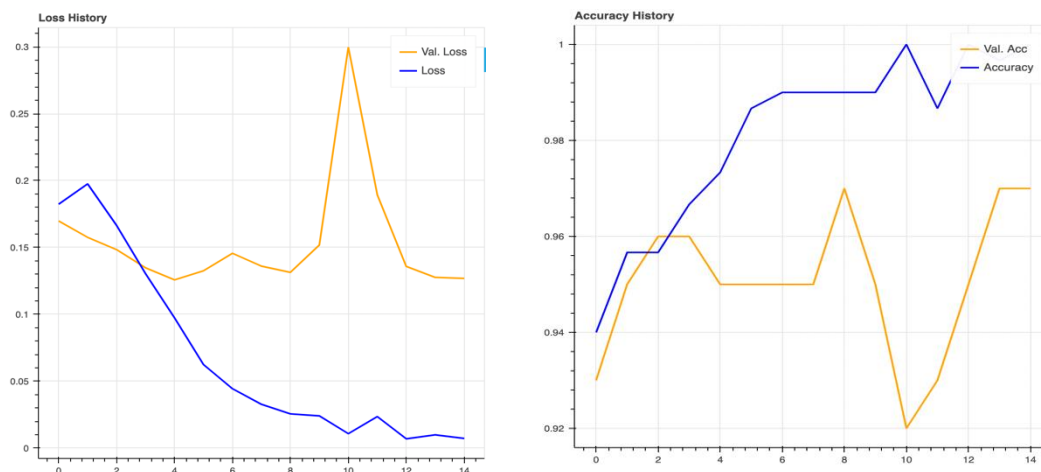


Gambar 3. Proses LSTM bekerja [3]

Pada penerapannya research ini menggunakan google colab sebagai media untuk melakukan run training, validation, dan testing model. Proses tersebut memakan waktu kurang dari satu jam. Adapun output atau hasil akurasi dan loss dari model dapat diperhatikan pada gambar berikut, Pada gambar tersebut terlihat bahwa nilai tertinggi dari akurasi mendekati angka 100%, sedangkan pada loss-nya nilai tertinggi maksimal di 3%

4. HASIL DAN PEMBAHASAN

Berdasarkan hasil trining, validation, dan testing, dengan menggunakan LSTM pada objek analisis sentiment pada komentar nitizen terhadap tayangan suatu acara di stasiun televisi menjelaskan bahwa loss sebesar 12% dengan tingkat akurasi sebesar 97% seperti yang diperlihatkan pada gambar berikut :



5. KESIMPULAN DAN SARAN

Dengan perkembangan ilmu pengetahuan dan teknologi yang begitu cepat, diharapkan mampu membantu dan mempermudah kehidupan manusia dalam mengambil suatu keputusan. Sebagai salah satu contoh, dengan menggunakan model LSTM dapat dilakukan suatu Analisa terhadap opini masyarakat mengenai suatu acara di stasiun televisi apakah bermuatan positif atau negatif dengan tingkat akurasi 97%. Ada begitu banyak hal yang bisa dikembangkan dengan machine learning, mulai dari metode, objek, jumlah dataset bahkan pengembangan terhadap model itu sendiri.

DAFTAR PUSTAKA

Referensi Cetak:

Buku

Penulis. *Judul buku*. Lokasi Penerbit: Penerbit, tahun, halaman.

- [1] Abu Farha, I., & Magdy, W. (2021). A comparative study of effective approaches for Arabic sentiment analysis. *Information Processing and Management*, 58(2). <https://doi.org/10.1016/j.ipm.2020.102438>
- [2] J.E. Bourne. "Synthetic structure of industrial plastics," in *Plastics*, 2nd ed., vol. 3. J. Peters, Ed. New York: McGraw-Hill, 1964, pp.15-67.

Jurnal

Penulis. "Judul Artikel". *Nama Jurnal*, vol., halaman, tanggal/tahun, DOI.

- [3] Christopher S. Goldenstein, et. al. "Infrared laser-absorption sensing for combustion gases." *Progress in Energy and Combustion Science*, Volume 60, May 2017, Pages 132-176, <https://doi.org/10.1016/j.pecs.2016.12.002>.

Prosiding

Penulis. "Judul Artikel." in *Conference proceedings*, tahun, halaman.

- [4] D.B. Payne and H.G. Gunhold. "Digital sundials and broadband technology," in *Proc. IOOC-ECOC*, 1986, pp. 557-998.

Peraturan Pemerintah

- [5] Republik Indonesia. Undang-Undang Nomor 3 Tahun 2014 tentang Perindustrian.

Standar

[6] British Standards Institution. *B.S. 764*. London: British Standards Institution, 1990.

Dokumen Paten

Inventor."Nama/Judul." Negara dimana paten terdaftar. Nomor, tanggal.

[7] E.E. Rebecca. "Alternating current fed power supply." U.S. Patent 7 897 777, Nov. 3, 1987.

Gambar Teknik

[8] F. Afrinaldi. *Rangka Belt Conveyor*. [Gambar Teknik]. Universitas Andalas: Padang, 2005.

Referensi Elektronik:

Buku

Penulis. (Tahun, bulan tanggal). *Judul buku*. (Edisi). [On-line]. Volume(Nomor). Available: site/path/file [tanggal diakses].

[9] S. Calmer. (1999, June 1). Engineering and Art. (2nd edition). [On-line]. 27(3). Available: www.enggart.com/examples/students.html [May 21, 2003].

Web

Penulis. "Judul." Internet: complete URL, tanggal di-update [tanggal diakses].

[10] M. Duncan. "Engineering Concepts on Ice. Internet: www.iceengg.edu/staff.html, Oct. 25, 2000 [Nov. 29, 2003].

Sumber Lain:

Koran

Penulis. "Judul Artikel." Nama Koran (tanggal, tahun), bagian/liputan, halaman.

[11] B. Bart. "Going Faster." *Globe and Mail* (Oct. 14, 2002), sec. A p.1. "Telehealth in Alberta." *Toronto Star* (Nov. 12, 2003), sec. G pp. 1-3.

Disertasi/Tesis/Tugas Akhir

Penulis. "Judul Tesis." Level Lulusan, nama universitas, lokasi, tahun.

[12] S. Mack. "Desperate Optimism." M.A. thesis, University of Calgary, Canada, 2000.