



**Analisis Sentimen Komentar Pada Platform *Youtube* Kategori Vlog  
Menggunakan Algoritma *Long Short Term Memory* network (LSTM)**



*TUGAS AKHIR*

Nikko Wahyudi

41518010170

UNIVERSITAS  
**MERCU BUANA**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2022**



**Analisis Sentimen Komentar Pada Platform Youtube Kategori Vlog  
Menggunakan Algoritma Long Short Term Memory network (LSTM)**

*Tugas Akhir*

Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer

Oleh:

Nikko Wahyudi

41518010170

UNIVERSITAS  
MERCU BUANA

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2022

## LEMBAR PERNYATAAN ORISINALITAS

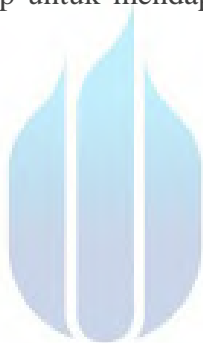
Yang bertanda tangan dibawah ini:

NIM : 41518010170

Nama : Nikko Wahyudi

Judul Tugas Akhir : Analisis Sentimen Komentar Pada Platform Youtube  
Kategori Vlog Menggunakan Algoritma Long Short Term  
Memory network (LSTM)

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan di dalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.



UNIVERSITAS  
MERCU BUANA

Jakarta, 12 Juli 2022



*Nikko*

Nikko Wahyudi

## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Nikko Wahyudi  
NIM : 41518010170  
Judul Tugas Akhir : Analisis Sentimen Komentar Pada Platform Youtube Kategori Vlog Menggunakan Algoritma Long Short Term Memory network (LSTM)

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non eksklusif** (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Non eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 12 Juli 2022



Nikko Wahyudi

## SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Nikko Wahyudi  
 NIM : 41518010170  
 Judul Tugas Akhir : Analisis Sentimen Komentar Pada Platform Youtube Kategori Vlog Menggunakan Algoritma Long Short Term Memory network (LSTM)

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis		Status	
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi		Diajukan	Ya
		Jurnal Nasional Terakreditasi	Ya		
		Jurnal International Tidak Bereputasi		Diterima	
		Jurnal International Bereputasi			
Disubmit/dipublikasikan di :	Nama Jurnal	: Jurnal Informatika Universitas Pamulang			
	ISSN	: 2541-1004			
	Link Jurnal	: <a href="http://openjournal.unpam.ac.id/index.php/informatika/index">http://openjournal.unpam.ac.id/index.php/informatika/index</a>			
	Link File Jurnal Jika Sudah di Publish	:			

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui

Jakarta, 12 Juli 2022

Dosen Pembimbing TA



Adi Hartanto, ST, M.Kom



Nikko Wahyudi



## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010170  
Nama : Nikko Wahyudi  
Judul Tugas Akhir : Analisis Sentimen Komentar Pada Platform Youtube Kategori Vlog Menggunakan Algoritma Long Short Term Memory network (LSTM)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 19 Agustus 2022



UNIVERSITAS  
MERCU BUANA  
(Ummiy Salamah, ST., MMSI)

## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010170  
Nama : Nikko Wahyudi  
Judul Tugas Akhir : Analisis Sentimen Komentar Pada Platform Youtube  
Kategori Vlog Menggunakan Algoritma Long Short  
Term Memory network (LSTM)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 4 Agustus 2022



(Sabar Rudiarto, M.Kom)



## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010170  
Nama : Nikko Wahyudi  
Judul Tugas Akhir : Analisis Sentimen Komentar Pada Platform Youtube  
Kategori Vlog Menggunakan Algoritma Long Short  
Term Memory network (LSTM)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 4 Agustus 2022



UNIVERSITAS  
MERCU BUANA

A handwritten signature in black ink, appearing to be 'Vina Ayumi'.

(Vina Ayumi, S.Kom., M.Kom)

## LEMBAR PENGESAHAN

NIM : 41518010170

Nama : Nikko Wahyudi

Judul Tugas Akhir : Analisis Sentimen Komentar Pada Platform Youtube  
Kategori Vlog Menggunakan Algoritma Long Short Term Memory  
network (LSTM)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 4 Agustus 2022

Menyetujui,



UNIVERSITAS  
MERCU BUANA

(Adi Hartanto, ST, M.Kom)


Dosen Pembimbing

Mengetahui,



(Wawan Guwawan, S.Kom, M

Koord. Tugas Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph.D., IPM.)

Ka. Prodi Teknik Informatika

## KATA PENGANTAR

Puji syukur saya panjatkan kehadirat Allah SWT yang telah memberikan segala rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan Tugas Akhir yang berjudul “Sentiment Analysis of Comments On *Youtube* Platform Using *Long Short Term Memory*network (LSTM) Algorithm” dalam jangka waktu yang sudah ditentukan. Tanpa pertolongan-NYA, mungkin saya tidak akan sanggup menyelesaikan dengan baik. Laporan Tugas Akhir ini dibuat sebagai syarat untuk LULUS sebagai sarjana Ilmu Komputer dari Universitas Mercu Buana. Penulis menyadari dalam pembuatan Tugas Akhir ini tidak terlepas dari bantuan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Keluarga saya, khususnya kedua orang tua saya yang telah menyemangati, mendukung dan selalu mendoakan yang terbaik untuk proses meraih gelar sarjana. Serta selalu percaya bahwa saya bisa menyelesaikan kuliah dengan baik.
2. Bapak Achmad Kodar, Drs. MT selaku Dosen Pembimbing Akademik yang telah memantau dan membantu studi saya hingga akhir.
3. Bapak Adi Hartanto, ST, M.Kom. selaku dosen Pembimbing Tugas Akhir yang telah memberikan masukan saat bimbingan dan meluangkan waktu sebagian besarnya untuk melakukan bimbingan dalam penyusunan tugas akhir ini hingga selesai.
4. Seluruh Dosen Program Studi Teknik Informatika yang sudah memberikan ilmu yang bermanfaat selama kuliah berlangsung. Memberi kesempatan untuk belajar, berkarya dan juga berkembang.
5. Seluruh Staff Administrasi dan Tata Usaha yang telah banyak membantu dan memberikan kemudahan, terima kasih atas semua pelayanan dan arahannya.
6. Sahabat dan kerabat, yang telah memberikan semangat, dukungan dan doa sehingga saya bisa melewati dan menyelesaikan Tugas Akhir ini dan mendapatkan gelar sarjana.

7. Pihak-pihak dan personal yang tidak dapat disebut satu per satu yang terlibat dalam pembuatan Tugas Akhir ini sehingga dapat selesai dengan baik.

Akhir kata, penulis berharap Tugas Akhir ini dapat menjadi referensi untuk dikembangkan lebih baik lagi kedepannya, karena penulis menyadari hasil Tugas Akhir ini masih jauh dari sempurna. Masih terdapat kekurangan dalam eksperimen,

cara penjelasan maupun kekeliruan penulisan. Untuk itu, kritik dan saran pembaca sangat dihargai dan diharapkan. Semoga Tugas Akhir ini dapat memberikan manfaat bagi para pembaca.

Jakarta, 12 Juli 2022

Penulis



## DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL .....	i
LEMBAR PERNYATAAN ORISINALITAS .....	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii	
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI.....	vi
LEMBAR PENGESAHAN .....	viii
ABSTRAK .....	x
ABSTRACT .....	xi
KATA PENGANTAR.....	xii
DAFTAR ISI.....	xiv
NASKAH JURNAL .....	1
KERTAS KERJA.....	8
BAB 1. LITERATUR REVIEW .....	9
BAB 2. ANALISIS DAN PERANCANGAN.....	18
BAB 3. SOURCE CODE .....	21
BAB 4. <i>DATASET</i> .....	25
BAB 5. TAHAPAN EKSPERIMEN.....	29
BAB 6. HASIL SEMUA EKSPERIMEN.....	34
DAFTAR PUSTAKA .....	39
LAMPIRAN DOKUMEN HAKI.....	42
LAMPIRAN KORESPONDENSI .....	44

## NASKAH JURNAL

### Analisis Sentimen Komentar Pada Platform Youtube Menggunakan Algoritma Long Short Term Memory network (LSTM)

Nikko Wahyudi<sup>1</sup> & Adi Hartanto<sup>2</sup>

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana  
[41518010170@student.mercubuana.ac.id](mailto:41518010170@student.mercubuana.ac.id)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana  
[adi.hartanto@mercubuana.ac.id](mailto:adi.hartanto@mercubuana.ac.id)

#### ABSTRAK

Sosial media merupakan tempat berbagi informasi yang sedang banyak digunakan pada zaman sekarang sehingga menimbulkan banyak reaksi sentimen. Youtube adalah salah satu penyedia media yang dipenuhi dengan beragam informasi yang mengakibatkan munculnya reaksi pengguna berupa sentimen baik dari positif maupun negatif. Dalam hal ini, diperlukan suatu alat untuk menganalisa sentimen para pengguna, dengan banyaknya pengguna yang semakin banyak tidak dimungkinkan untuk menganalisa secara manual yang akan menggunakan waktu terlalu lama, salah satu metode analisis sentimen yang bisa digunakan adalah Long Short Term Memory (LSTM). Berdasarkan pengujian yang dilakukan terdapat parameter terbaik untuk membangun metode Long Short Term Memory yaitu Dropout sebesar 0.8, Epoch dengan jumlah 20 serta batch size dengan jumlah 32 yang menghasilkan akurasi sebesar 86%.

#### Kata kunci :

Analisis Sentimen; Long Short Term Memory; Sosial Media; Youtube

#### ABSTRACT

Social media is a place to share information that is being widely used today, causing many sentiment reactions. Youtube is one of the media providers that is filled with a variety of information that results in the emergence of user reactions in the form of sentiments both from positive and negative. In this case, a tool is needed to analyze the sentiment of users, with an increasing number of users it is not possible to analyze manually which will take too long, one of the sentiment analysis methods that can be used is Long Short Term Memory (LSTM). Based on the tests carried out, there are the best parameters to build the Long Short Term Memory method, namely Dropout of 0.8, Epoch with number of 20 and batch size with a total of 32 which results in an accuracy of 86%.

#### Keywords :

Sentiment Analysis; Long Short Term Memory; Social Media; Youtube

#### PENDAHULUAN

Pada era modern ini semakin berkembangnya teknologi di segala bidang mengakibatkan mempermudah manusia untuk melakukan segala pekerjaan pada bidangnya masing - masing. Tidak terkecuali teknologi informasi yang sangat menonjol dalam kemajuan teknologinya, bandingkan pada zaman dahulu media informasi hanya bisa didapat melalui teks yang ditulis di sebuah kertas sedangkan di zaman sekarang informasi dapat diperoleh dengan teks, gambar, suara dan

media lainnya bahkan informasi ini dapat diperoleh di manapun dan kapanpun dengan mudahnya. Informasi yang telah diberikan kepada orang ini juga tidak hanya bisa dinikmati saja tetapi pembaca/penonton dapat berinteraksi dengan para kreator yang biasanya menggunakan komentar. Salah satu media informasi yang menyediakan interaksi antara kedua belah pihak yang terbesar di dunia ini yang dapat memberikan informasi secara jelas dan bisa menggunakan segala media yang ada sekarang adalah *youtube.com*.

Komentar adalah sebuah tanggapan dari seorang pembaca/penonton setelah melihat konten yang diberikan oleh kreator baik itu bertanya, memberikan saran, atau hanya sekedar mengekspresikan diri mereka. Dengan begitu para kreator dapat mengetahui tanggapan yang diberikan oleh para penonton. Tanggapan ini dapat digunakan oleh para kreator untuk mengevaluasi konten agar dapat menjadi lebih baik.

Media memiliki kekuatan untuk mempengaruhi pandangan masyarakat dalam proses pembentukan opini atau sudut pandangnya. Namun apa yang ditampilkan media bukanlah sesuatu yang tetap dan dimaknai secara utuh oleh khalayak. Khalayak juga memiliki pandangan tersendiri dalam memaknai teks media. Maka tidak heran jika pemaknaan khalayak terhadap berita yang sama itu menjadi berbeda-beda [1].

Salah satu cara para konten kreator dapat mengembangkan konten adalah dari reaksi para penonton. Maka dari itu sangat penting untuk mencari tahu reaksi yang diberikan penonton salah satunya dengan menganalisis komentar. Dalam dunia komputer analisis ini dikenal sebagai analisis sentimen (*sentiment analysis*). Analisis sentimen ini akan menganalisis komentar para penonton secara otomatis.

Analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen (Huang, 2009) [2].

Penelitian yang dilakukan oleh Jovita Nurvania, Jondri dan Kemas Muslim Lhaksamana mengklasifikasikan ulasan pengunjung tentang pengaruh COVID-19 terhadap tempat wisata di Bali dari Tripadvisor menggunakan metode *Long Short-Term Memory* (LSTM). Sebelum diproses dengan LSTM, setiap teks pada ulasan akan di vektorisasi dengan word2vec. Hasil pengujian pada model yang dibangun didapatkan nilai akurasi sebesar 71,67%. [3] Pada penelitian tersebut pada tahap labelling masih dilakukan secara manual.

Berdasarkan penelitian terkait yang telah dilakukan sebelumnya, penelitian ini memiliki perbedaan yaitu pada tahap labelling yang dilakukan dengan cara otomatis menggunakan vader dan dataset yang digunakan adalah komentar dari *youtube*.

## TINJAUAN PUSTAKA

### 2.1 Sentiment Analysis

*Sentiment analysis* (analisis sentimen) atau sering disebut juga dengan *opinion mining* (penambangan opini) adalah studi komputasi untuk mengenali dan mengekspresikan opini, sentimen, evaluasi, sikap, emosi, subjektivitas, penilaian atau pandangan yang terdapat dalam suatu teks (Liu 2008). Analisis sentimen dilakukan guna menilai opini dan kecenderungan sebuah opini terhadap suatu topik baik negatif maupun positif (Rozi, Pramono dan Dahlan, 2012). Analisis sentimen dapat diterapkan pada opini semua bidang seperti ekonomi, politik, sosial dan hukum. Media sosial Twitter ini membuka jendela bagi para peneliti untuk mempelajari emosi, suasana hati, dan pendapat publik melalui analisis sentimen (Qiu, Lin dan Shuai, 2019) [4]. Analisis sentimen juga bisa ditafsirkan sebagai informasi tekstual umumnya dibagi menjadi fakta (objektif) dan opini (subyektif) informasi [5].

### 2.2 Situs Youtube

*Youtube* merupakan situs *video sharing* yang sangat populer dikarenakan konten yang tersedia pada *youtube* sangatlah beragam serta bisa dipastikan konten yang terdapat pada *youtube* lebih banyak dibandingkan konten yang disediakan oleh stasiun tv bahkan untuk layanan televisi berbayar sekalipun. Bahkan pada saat ini stasiun televisi sudah ada beberapa yang mengunggah konten yang seharusnya ada di televisi juga diunggah ke *youtube* karena memang pasar *youtube* sangatlah besar. Karena *youtube* ini berbasis digital sehingga sangat mudah untuk dinikmati oleh kalangan apapun apalagi dengan mudahnya internet sekarang ini semakin menjadikan *youtube* penyedia konten *video sharing* yang terbesar saat ini.

### 2.3 LSTM

*Long Short Term Memory* (LSTM) merupakan salah satu jenis pengklasifikasian teks yang berevolusi dari arsitektur RNN, yang pertama kali diperkenalkan oleh Hochreiter dan Schmidhuber. [3] Pada LSTM, nilai error saat

proses backpropagation dilakukan melalui waktu dan layer akan disimpan, sehingga memungkinkan LSTM untuk menyimpan.

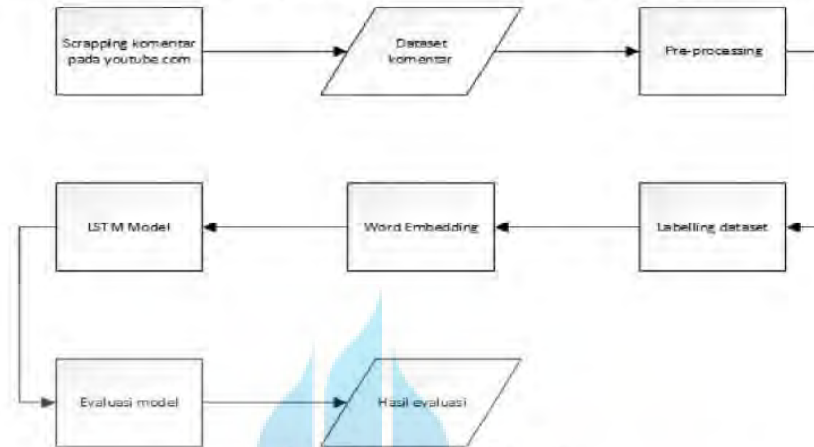
2.4 Dropout

Dropout digunakan agar model tidak tidak mengalami kondisi *overfitting*. Dropout ini bekerja dengan cara menonaktifkan beberapa *neuron*, sehingga membuat *neuron* akan bekerja sendiri dan tidak bergantung pada neuron lainnya. Rentang nilai pada dropout ini

*stopwords*, *tokenizing*, *case folding* dan *stemming*. Tujuan dari tahapan preprocessing ini adalah agar dataset yang nantinya akan dilakukan pengolahan data akan lebih mudah diolah.

1. Cleansing

Teks hasil scraping dari *youtube* selanjutnya di tahap ini akan diproses pada tahap *cleansing* untuk menghilangkan user, tagar, url, emoji, dan *Spelling correction*.



Gambar 1. Tahapan Penelitian

adalah 0.0 – 1.0

**METODE PENELITIAN**

Alur pada penelitian ini dapat dilihat pada gambar nomor.

3.1 Pengumpulan data

Data yang akan digunakan untuk penelitian ini berasal dari seorang *youtuber* dengan nama channel AH (Atta Halilintar) serta judul konten yang dituju adalah ATTA Jemput Fateh Sekolah PERTAMA Kali pake MOBIL i8.RUSUH!. Dataset akan memiliki beberapa kolom yaitu *Name, Comment, Time, Likes, Reply Count, Reply, Published, Updated* dengan format *CSV* yang akan diproses di tahapan selanjutnya.

Tabel 1. Contoh data

Kagum banget sama bang atta ramah banget sama semua orang sama sekali ngga sombong walaupun udah tinggi ☺ keep humble bang atta ☺☺
--

3.2 Preprocessing

Pada tahap preprocessing ini data yang telah berbentuk *CSV* akan melalui beberapa tahap yaitu *cleansing*, *spelling correction*,

*Cleansing* bertujuan agar meringankan proses saat komputasi serta membuat efisien.

2. Case folding

Di *case folding* ini data teks akan diubah menjadi lebih sederhana lagi untuk mempermudah kinerja model. *Case folding* ini bertujuan agar data teks yang mempunyai huruf kapital akan diubah menjadi huruf kecil agar menjadi lebih sederhana.

3. Tokenizing

Untuk mempermudah proses analisis *tokenizing* juga perlu, dalam proses ini simbol, angka dan *whitespace* akan dihilangkan karena teks seperti itu tidak berguna pada saat dalam proses pemodelan.

4. Stopword

Pada proses ini kata yang tidak memiliki makna akan dihapus serta ada tambahan kata kustom yang ditambahkan agar data menjadi lebih bersih.

5. Stemming

Proses ini adalah tahapan di mana memfilter kata dari hasil *token*



sebelumnya. Yang *difilter* pada saat tahapan ini adalah kata penghubung yang terdapat pada data teks, agar teks berubah menjadi ke kata dasar.

Tabel 2. Hasil preprocessing

Sebelum	Sesudah
Bang atta jago juga main bolanya 🍀❤️	bang atta jago main bola
Keren2 Konten Kali Ini !	keren konten kali ini

### 3.3 Translate

Setelah dataset melalui tahap *cleansing* maka selanjutnya dataset akan di *translate* ke bahasa Inggris dalam kasus dataset yang digunakan bahasa Indonesia ke bahasa Inggris. Tujuan dari *translate* ini adalah agar dataset dapat memenuhi persyaratan untuk menggunakan metode *lexicon based* dengan *vader*.

### 3.4 Labeling Dataset

Di tahap *labeling* dataset ini data yang sudah melewati tahap *cleansing* akan diberi label menggunakan metode *lexicon based* dan algoritma *vader* dengan *output*  $> 0$  = negatif,  $= 0$  = neutral dan  $< 0$  = positif. Dari keseluruhan data sebanyak 11501 yang ada didapat bahwa komentar positif 5307 data, komentar neutral 5058 data dan komentar negatif 1136. Pada penelitian ini yang digunakan adalah data dengan label positif dan negatif saja, karena pada label neutral kebanyakan mengandung kata maupun kalimat yang tidak bermakna.

Tabel 3. Labeling

Comment	Compound	Label
abang atta sombong	-0.0498	negatif
kak paling hidup kurang asik bang atta	f	f
bang lu ganteng banget	0.6269	positif

### 3.5 UnderSampling

Pada keseluruhan data hasil dari *labeling* didapat bahwa komentar positif 5307 data komentar negatif sebanyak 1136 oleh karena itu dataset dianggap tidak seimbang. Maka akan dilakukan *undersampling* sehingga data dengan label positif dan negatif akan memiliki rasio 50:50. Data akan berjumlah sama yaitu 1100 data dengan label positif dan 1100 data dengan label negatif.

### 3.6 Word Embedding

Setelah itu dataset akan menuju tahap *word embedding* yang bertujuan untuk kata unik agar diubah menjadi dalam bentuk *vector* dengan menggunakan keras. Pengubahan kata menjadi *vector* ini agar dapat diproses melalui komputasi dalam tahap berikutnya.

### 3.7 LSTM Model

*Long Short-Term Memory* (LSTM) adalah salah satu variasi dari *Recurrent Neural Network* yang dibuat untuk menghindari masalah ketergantungan jangka Panjang pada *Recurrent Neural Network* (RNN). LSTM dapat mengingat informasi jangka panjang. Pada RNN perulangan jaringan hanya menggunakan satu layer sederhana [6]. LSTM ini mempunyai model yg sama dengan RNN dikarenakan LSTM ini sendiri hanya salah satu metode yang terdapat pada RNN hanya saja perbedaan antara RNN dan LSTM ini terdapat pada strukturnya, pada LSTM memiliki layer pada setiap perulangan yang dilakukan.

LSTM memiliki 3 jenis *gates* di antaranya adalah *forget gate*, *input gate* dan *output gate*.

#### 1. Forget gate

Pada gerbang ini nilai output sebelumnya dengan input saat ini digabung lalu melewati fungsi aktivasi *sigmoid*. Gerbang inilah yang menentukan apakah informasi sebelumnya akan dilupakan atau tidak. Kemudian informasi ini dilanjutkan ke *memory cell* atau *cell state*.

$$f_t = \sigma(W_f \times [x_t + h_{t-1}] + b_f)$$

Keterangan :

- $f_t$  : forget gate
- $\sigma$  : sigmoid activation function
- $W_f$  : weights forget gate
- $x_t$  : input cell
- $h_{t-1}$  : output cell previously
- $b_f$  : bias forget gate

#### 2. Input Gate (it)

Pada gerbang ini nilai output sebelumnya dengan input saat ini digabung, lalu ada dua fungsi aktivasi yang akan dilewatinya. Jalur satu melewati fungsi aktivasi *sigmoid* untuk nilai *input*, jalur lainnya melewati fungsi aktivasi *tanh* untuk nilai *candidate memory cell*. [3]

$$i_t = \sigma(W_i \times [x_t + h_{t-1}] + b_i)$$

$$\tilde{C}^t = \tanh(WC \times [xt + ht-1] + bC)$$

Keterangan :

- $I^t$  : input gate
- $W_i$  : weights input
- $b_i$  : bias input gate
- $\tilde{C}^t$  : candidate
- $\tanh$  : tanh activation function
- $WC$  : weights candidate
- $bC$  : bias candidate

### 3. Output Gate

Gerbang ini menghasilkan nilai *output*, dimana nilai ini berasal dari gabungan nilai sebelumnya dengan nilai saat ini yang telah melalui fungsi aktivasi *sigmoid*. [3]

$$ot = \sigma(Wo \times [xt + ht-1] + bo)$$

Keterangan :

- $ot$  : output gate
- $Wo$  : weights output gate
- $bo$  : bias output gate

### 3.8 Evaluasi Model

Evaluasi disini untuk mengetahui tingkat akurasi, *loss*, *recall*, *precision* dan *f1 score* dari model yang sudah dibuat sehingga akan terlihat seberapa baik model pada penelitian ini.

Berikut merupakan rumus yang digunakan untuk mengetahui tingkat akurasi, *loss*, *recall*, *precision* dan *f1 score*

- *Accuracy*

*Accuracy* adalah perhitungan seberapa tepat klasifikasi yang telah dibangun, sesuai dengan target yang ada. [3]

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

- *Precision*

*Precision* adalah perhitungan keakuratan antara data target dengan hasil prediksi dari model. [3]

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

- *Recall*

*Recall* adalah perhitungan yang menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. [3]

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

- *F1-Score*

*F1-score* adalah perhitungan yang menggambarkan perbandingan antara *precision* dan *recall*. Jika nilai FN dan FP tidak mendekati sebaiknya digunakan *f1-score* dibandingkan nilai akurasi. [3]

$$Recall = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\%$$

Keterangan :

- *True positive* (TP) : Prediksi yang bernilai positif dan benar sesuai target.
- *True negative* (TN) : Prediksi yang bernilai negatif dan benar sesuai target.
- *False positive* (FP) : Prediksi yang bernilai positif dan salah tidak sesuai target.
- *False negative* (FN) : Prediksi yang bernilai negatif dan salah tidak sesuai target.

## HASIL DAN PEMBAHASAN

### 4.1 Pengujian Model

Pemilihan model adalah tahapan dimana untuk memilih model yang terbaik dari tes uji coba yang ada untuk dimasukkan kedalam sistem. Pada tahapan model ini dataset yang digunakan sebanyak 2200 data dari

Tabel 4. Pengujian Dropout

Dropo ut	Epo c h	Data training		Data validation	
		Acc %	Los s	Acc %	Loss
0.2	20	98%	0.06	83%	0.73
0.4		94%	0.13	85%	0.41
0.6		93%	0.17	86%	0.37
0.8		96%	0.12	86%	0.41

hasil undersampling dengan data label positif dan negatif 50:50. Setelah itu data komentar tersebut akan di bagi sebesar 70% untuk data *training* dan 30% untuk data *testing*.

Pada metode *Long Short Term Memory*, akan dilakukan beberapa pengujian untuk mendapat parameter yang terbaik di antaranya

pengujian jumlah *epoch* dan pengujian jumlah *dropout*.

Pada pengujian pertama, dilakukan pengujian *dropout* dengan *dropout* sebesar 0.2, 0.4, 0.6, 0.8 yang dapat dilihat pada tabel 6 Pengujian dilakukan dengan *epoch* sebesar 20. Pada pengujian *dropout* dapat dilihat bahwa jumlah *dropout* dengan akurasi yang terbaik terdapat pada *dropout* sebesar 0.8 dengan akurasi sebesar 86% dan *loss* hanya 0.41, dengan pengujian tersebut juga menunjukkan bahwa semakin besar *dropout* maka hasil dari akurasi juga semakin tinggi dengan hasil yang semakin meningkat seiring ditambahkan jumlah *dropout*. Hal ini dapat terjadi karena akan membuat model menjadi sederhana dan tidak mengalami *overfitting*, dimana model tidak akan *mentraining* data terlalu kompleks sehingga tidak memberatkan model.

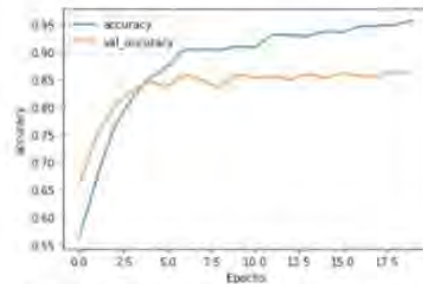
Pada tes yang kedua, akan dilakukan pengujian pada jumlah *epoch* dengan jumlah 12, 14, 16, 18, 20 yang dapat dilihat pada tabel 5 pengujian dilakukan dengan *dropout* sebesar 0.8 berdasarkan hasil parameter *dropout* yang terbaik pada pengujian sebelumnya. Dapat dilihat bahwa pengujian *epoch* terbaik terdapat pada *epoch* dengan jumlah 20 dengan akurasi sebesar 88% dan *loss* sebesar 0.3.

Berdasarkan dari hasil pengujian yang telah dilakukan pada pengujian *dropout* dan pengujian *epoch* didapatkan bahwa hasil maksimal dari masing-masing parameter dari model yang telah melakukan *training* menggunakan data train yang dilihat dari data *validation* pada akurasi tertinggi dan *loss* terendah saat menggunakan *dropout* sebesar 0.8 dan *epoch* sebanyak 20. Dapat dilihat pada tabel 5 perbandingan dari data *training* dan data *validation* menunjukkan bahwa akurasi data train sebesar 96% dan *loss* 0.12 sedangkan pada data *validation* mendapatkan akurasi sebesar 86% dan *loss* 0.41. model LSTM yang dibangun sudah cukup baik terlihat juga dari graphic pada gambar 2 untuk melihat perkembangan grafik akurasi dan pada gambar 3 untuk melihat tingkat *loss*.

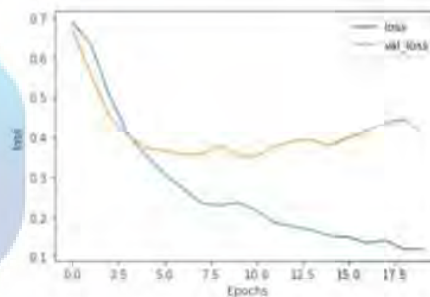
Berdasarkan penelitian yang telah dibuat dengan model LSTM, model tidak akan memprediksi data baru yang tidak ada pada dataset yang telah direpresentasikan kedalam *word index*. Oleh karena itu model ini memerlukan dataset yang lebih banyak agar dapat memvariasi fitur kata dalam *word index* lebih banyak lagi.

Tabel 5. Pengujian epoch

Dropout	Epoch	Data training		Data validation	
		Acc %	Loss	Acc %	Loss
0.8	12	93%	0.18	86%	0.38
	14	93%	0.18	85%	0.4
	16	94%	0.15	86%	0.35
	18	95%	0.14	86%	0.44
	20	96%	0.12	86%	0.41



Gambar 2. grafik akurasi training dan validation



Gambar 3. grafik loss training dan validation

Selanjutnya akan dilakukan pengujian pada masing-masing kelas sentimen untuk mengetahui seberapa bagus model LSTM yang telah dibangun dalam keakuratan untuk menilai setiap sentimen.

Parameter yang digunakan adalah parameter dengan hasil terbaik pada pengujian sebelumnya yaitu *dropout* sebesar 0.8 dan *epoch* dengan jumlah 20 untuk menghasilkan tingkat keakuratan yang maksimal. Dengan parameter terbaik tersebut didapat bahwa tingkat akurasi berdasarkan data *validation* untuk sentimen positif sebesar 86% sedangkan untuk sentimen negatif mendapatkan akurasi sebesar 86%. Hal ini menunjukkan model LSTM yang dibangun dari hasil akurasi di setiap sentimennya cukup baik.

Tabel 6. Akurasi data validation setiap kelas

Kelas	Acc %
positif	86 %
negatif	86 %

Tabel 7. Performansi model

Acc %	Precision %	Recall %	F1-score %
86%	86%	86%	86%

### KESIMPULAN

Berdasarkan hasil dari penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa penerapan algoritma *Long Short Term Memory* pada komentar di *youtube* secara otomatis menghasilkan akurasi yang cukup baik dengan tingkat akurasi sebesar 86% pada saat menggunakan *data training* sebanyak 70% (1540 komentar) dan *data testing* sebanyak 30% (660 komentar) dengan jumlah total data 2200 dari hasil *undersampling*. Model juga bisa melakukan analisis sentimen dengan mengkategorikan ke 2 jenis kelas yaitu positif dan negatif. Model pada penelitian ini menggunakan parameter pada performa terbaik di *dropout* sebesar 0.8 dan *epoch* sebanyak 20.

### DAFTAR PUSTAKA

- [1] I. F. Dewanti, "Analisis Isi Komentar Pemberitaan Pada Portal Berita Republika Online (Studi Analisis Isi Komentar Pemberitaan tentang Pro Kontra Menteri Susi Pudjiastuti pada Portal Berita Republika Online Periode 27 Oktober 2014 – 13 November 2014)," no. November, pp. 1–20, 2014.
- [2] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *INTEGER J. Inf. Technol.*, vol. 1, no. 1, 2017.
- [3] J. Nurvania and K. M. Lhaksamana, "Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode Long Short-Term Memory ( LSTM )," *e-Proceeding Eng.*, vol. 8, no. 4, 2021.
- [4] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, 2021, doi: 10.25126/jtiik.0813944.
- [5] B. Pratama *et al.*, "Sentiment Analysis of the Indonesian Police Mobile Brigade Corps Based on Twitter Posts Using the SVM and NB Methods," in *Journal of Physics: Conference Series*, 2019, vol. 1201, no. 1, doi: 10.1088/1742-6596/1201/1/012038.
- [6] I. F. Rozi, V. N. Wijayaningrum, and N. Khozin, "KLASIFIKASI TEKS LAPORAN MASYARAKAT PADA SITUS LAPOR! MENGGUNAKAN RECURRENT NEURAL NETWORK," *SISTEMASI*, vol. 9, no. 3, 2020, doi: 10.32520/stmsi.v9i3.977.
- [7] C. K. N. Papatungan and A. Jacobus, "Sentiment Analysis of Social Media Users Using Long-Short Term Memory Method," *J. Tek. Elektro dan Komput.*, vol. 10, no. 2, 2021.

## KERTAS KERJA

### Ringkasan

Kertas kerja ini berisi tentang kelengkapan material dari artikel jurnal dengan judul “Analisis Sentimen Komentar Pada Platform Youtube Kategori Vlog Menggunakan Algoritma Long Short Term Memory network (LSTM)”. Seluruh hasil penelitian Tugas Akhir yang tidak dimasukkan kedalam artikel jurnal. Pada kertas kerja ini disajikan: *literatur review*, analisis perancangan, *source code*, *dataset* yang digunakan, tahapan eksperimen dan seluruh hasil eksperimen.

- Bagian 1: *Literature Review* menjabarkan mengenai beberapa jurnal yang terkait dengan penelitian.
- Bagian 2: Analisis dan Perancangan menjabarkan analisis masalah dan analisis model.
- Bagian 3: *Source Code* menjabarkan kumpulan kode di setiap prosesnya mulai dari Membaca data, Preprocessing, Implementasi model, sampai pengujian dan evaluasi.
- Bagian 4: *Dataset* melampirkan pengambilan *dataset*.
- Bagian 5: Tahapan eksperimen berisi pengumpulan data, *Preprocessing*, Implementasi model, dan evaluasi model.
- Bagian 6: Hasil implementasi eksperimen secara keseluruhan yang mencakup jawaban dari skenario percobaan pengujian yang sudah disebutkan pada bagian 5. Bagian ini berguna untuk menemukan model prediksi terbaik yang dilihat dari beberapa pengujian berdasarkan nilai dropout serta jumlah *epoch*. Kemudian, hasil yang didapat digunakan sebagai pemodelan untuk sentiment pada komentar *youtube* menggunakan algoritma Long Short Term Memory.