

# Analisis Sentimen Pada Twitter Terhadap Isu Penundaan Pemilu 2024 Dengan Membandingkan Metode *Long Short-Term Memory* Dan *Naïve Bayes Classifier*

Fahmi Mahmuji Cholis<sup>1, a)</sup> Muchammad Chandra Cahyo Utomo<sup>2, b)</sup> Nisa Rizqiya Fadhliana<sup>3, c)</sup>

<sup>1,2,3</sup>Institut Teknologi Kalimantan

<sup>a)</sup>Penulis korespondensi: 11181026@student.itk.ac.id

<sup>b)</sup>ccahyo@lecturer.itk.ac.id

<sup>c)</sup>nisafadhliana@lecturer.itk.ac.id

**Abstrak.** Negara Indonesia memiliki sistem pemerintahan yang berbentuk sistem demokrasi, oleh karena itu pada sistem demokrasi untuk memilih orang yang akan menjadi pejabat politik diperlukan adanya sebuah proses pemilihan. Proses tersebut dikenal dengan Pemilu (Pemilihan Umum). Menjelang pemilu pada 2024 muncul isu wacana penundaan Pemilu 2024. Di Indonesia, pengguna Twitter pada tahun 2022 sebanyak 18,45 juta. Dilihat dari banyaknya pengguna Twitter dengan munculnya isu penundaan pemilu, membuat banyak orang menyampaikan opininya di Twitter terkait isu tersebut. Penelitian ini untuk dapat mengetahui opini masyarakat secara umum pada Twitter terhadap isu penundaan Pemilu 2024 dengan membandingkan metode *Long Short-Term Memory* dan *Naïve Bayes Classifier*. Penelitian dilakukan dengan tahap *crawling*, *pre-processing*, pelabelan data, pembagian data *train* dan data *test* dengan perbandingan 9:1. Pada metode *Long Short-Term Memory* memperoleh nilai *accuracy* sebesar 92%, *precision* untuk kelas negatif sebesar 92% dan kelas positif sebesar 92%, *Recall* untuk kelas negatif sebesar 92% dan kelas positif sebesar 92%, *F1-Score* untuk kelas negatif sebesar 92% dan kelas positif sebesar 92%. Pada metode *Naïve Bayes Classifier* memperoleh nilai *accuracy* sebesar 80%, *precision* untuk kelas negatif sebesar 83% dan kelas positif sebesar 77%, *Recall* untuk kelas negatif sebesar 79% dan kelas positif sebesar 82%, *F1-Score* untuk kelas negatif sebesar 81% dan kelas positif sebesar 80%. Hasil analisis sentimen pada Twitter terhadap isu penundaan Pemilu 2024 dengan metode *Long Short-Term Memory* didapat sentimen positif sebesar 52.9% dan sentimen negatif sebesar 47.1%, sedangkan dengan metode *Naïve Bayes Classifier* didapat sentimen positif sebesar 32.5% dan sentimen negatif sebesar 67.5%.

## PENDAHULUAN

Negara Indonesia memiliki sistem pemerintahan yang berbentuk sistem demokrasi, oleh karena itu pada sistem demokrasi untuk memilih orang yang akan menjadi pejabat politik diperlukan adanya sebuah proses pemilihan. Proses tersebut dikenal dengan Pemilu (Pemilihan Umum). Di Indonesia Pemilu dilaksanakan sekali dalam kurun waktu lima tahun [1]. Menjelang dilakukannya pemilu pada 2024 muncul isu wacana penundaan Pemilu 2024. Pada awal tahun 2022, wacana mengenai penundaan Pemilu 2024 menjadi perbincangan hangat. Usul terkait penundaan pemilu disuarakan pada Januari 2022 oleh Menteri Investasi/Kepala BKPM Bahlil Lahadalia. Beliau menyampaikan pendapat dari para usahawan mengenai pelaksanaan Pemilu 2024. Beliau mengakui jika penundaan pemilu diharapkan oleh para pengusaha [8].

Perkembangan teknologi informasi memberi dampak semakin cepat dan mudahnya berbagi komunikasi dan informasi. Hal ini bisa dilihat dari adanya perkembangan media sosial. Media komunikasi dan informasi salah satunya yang dipakai oleh banyak orang yaitu Twitter. Di Indonesia, pengguna Twitter pada tahun 2022 sebanyak 18,45 juta, sehingga membuat Indonesia berada di peringkat kelima dunia [9]. Twitter memungkinkan penggunanya untuk dapat menyampaikan dan membahas berbagai topik atau isu yang sedang terjadi. Pengguna Twitter bisa menyampaikan opini pada *tweet* sesuai dengan keinginannya. *Tweet* dapat berupa saran, pendapat, ataupun kritikan mengenai topik yang ada [2].

Isu penundaan pemilu membuat banyak orang menyampaikan opininya pada Twitter terkait isu tersebut. Opini yang disampaikan dapat dikelompokkan menjadi opini yang bernilai positif ataupun negatif. Data opini bisa menjadi suatu informasi yang bermanfaat jika data tersebut dikumpulkan dan diolah, dengan dilakukannya analisis sentimen.

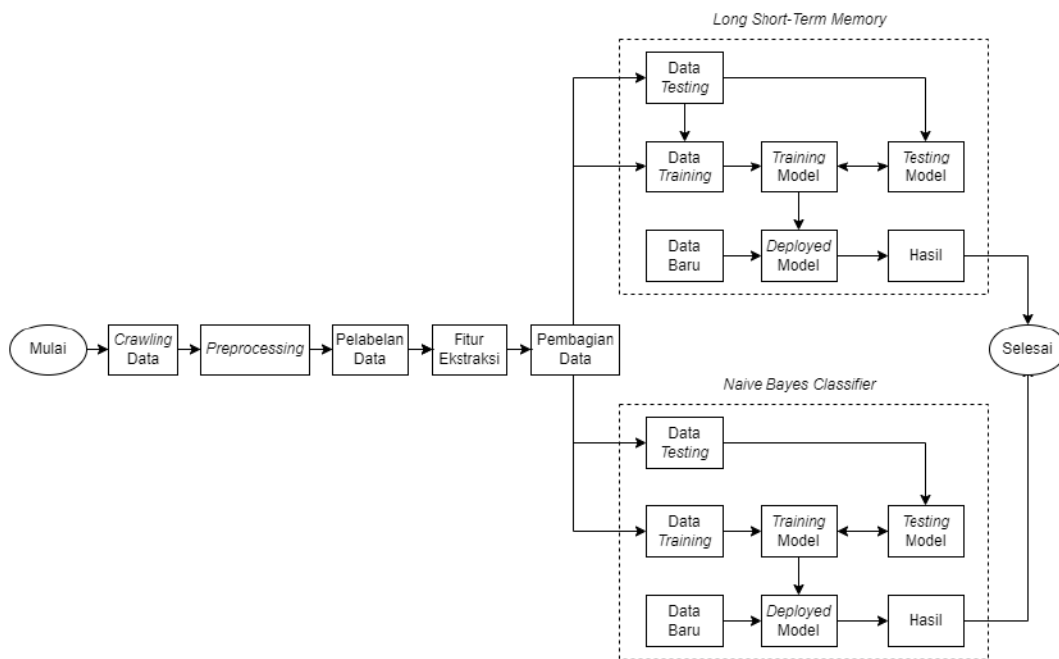
Analisis sentimen diterapkan guna mengklasifikasikan data, sehingga didapatkan data bernilai positif dan juga negatif dari hasil klasifikasi [1].

Dalam pengklasifikasian data teks terdapat beberapa metode. Pada penelitian ini menggunakan metode LSTM (*Long Short-Term Memory*) dan NBC (*Naïve Bayes Classifier*). LSTM adalah salah satu model dari *deep learning*, model ini bisa dipakai dalam melakukan pengklasifikasian data. LSTM biasanya digunakan untuk pemrosesan data sekuensial semacam data teks, video, dan suara [3]. NBC adalah algoritma yang bisa digunakan untuk klasifikasi sentimen dengan mencari nilai probabilitas tertinggi [4]. Penelitian ini dilakukan untuk dapat mengetahui opini masyarakat secara umum pada Twitter terhadap isu penundaan Pemilu 2024 dengan membandingkan metode *Long Short-Term Memory* dan *Naïve Bayes Classifier*.

## METODOLOGI PENELITIAN

### Diagram Alir Penelitian

Tahapan-tahapan penelitian yang dilakukan penulis dapat dilihat pada **GAMBAR 1**.



**GAMBAR 1.** Diagram alir penelitian

Pada **GAMBAR 1** penelitian ini dimulai dengan tahap *crawling* data, *crawling* dilakukan pada media sosial Twitter dengan kata kunci “penundaan pemilu” dan “tunda pemilu”. Kemudian data hasil *crawling* dilakukan tahap *pre-processing* berupa *cleaning*, *case folding*, *tokenization*, normalisasi, *stopword removal*, *stemming*, penghapusan data duplikat dan data kosong. Kemudian hasil dari *pre-processing* dilakukan tahap pelabelan data otomatis dengan menggunakan kamus *Lexicon-based*. Kemudian tahap fitur ekstraksi dengan mengubah data berbentuk kata menjadi data berbentuk vektor. Kemudian tahap pembagian data dilakukan untuk memecah data menjadi 2 kelompok data yaitu data *train* dan data *test*. Kemudian hasil dari pembagian data digunakan untuk tahap *training* dan *testing* model. Pada tahap *training* model digunakan dua model yaitu *training* model LSTM (*Long Short-Term Memory*) dan *training* model NBC (*Naïve Bayes Classifier*). Kemudian dari masing-masing model tersebut dilakukan evaluasi performa, kemudian klasifikasi model menggunakan data baru untuk dianalisis. Kemudian setelah didapatkan klasifikasi dari kedua model maka penelitian selesai.

## ***Crawling Data***

Pada tahap ini dilakukan pengambilan data pada Twitter menggunakan API. Langkah pertama yang dilakukan yaitu mendaftar pada Twitter *developer* untuk mendapatkan *access token*, *access token secret*, *api key*, *api key secret*. Kemudian pada proses *crawling* digunakan *library* yang bernama *Tweepy*. Data yang diambil dengan menggunakan kata kunci “penundaan pemilu” dan “tunda pemilu”. Selanjutnya data disimpan dalam bentuk file dengan format .csv.

## ***Pre-processing***

*Pre-processing* adalah tahap pembersihan *tweet* yang didapat di Twitter. Tahap pembersihan dilakukan dengan menghilangkan kondisi yang tidak dibutuhkan guna memperoleh data yang bermutu dan sesuai dari keperluan [3]. Tujuan *pre-processing* data yaitu untuk mentransformasikan data ke bentuk format yang dapat lebih mudah untuk dipelajari [1].

Pada penelitian yang dilakukan oleh [5] *pre-processing* dilaksanakan dengan proses berikut:

1. *Cleaning*, yaitu tahap penghapusan karakter yang memang tidak dibutuhkan seperti *username*, *hashtag*, *mention*, tanda baca, URL, angka, dan emotikon yang terdapat di isi *tweet*.
2. *Case folding*, yaitu tahap penyesuaian huruf pada teks, untuk diubah huruf besar ke huruf kecil.
3. Tokenization, yaitu tahap pemecahan suatu kalimat ke bentuk token dari setiap kata yang tersusun.
4. Normalisasi, yaitu tahap penyesuaian kata yang tidak terdapat pada KBBI.
5. *Stopword removal*, yaitu tahap penghapusan kata yang tidak mempunyai nilai, sehingga hanya kata yang mempunyai nilai saja.
6. *Stemming*, yaitu tahap perubahan kata ke dalam bentuk kata dasar.

## ***Pelabelan Data***

Pada tahap ini dilakukan pelabelan data yang sudah didapat dari hasil *crawling*. Pelabelan dilakukan secara otomatis dengan menggunakan *Lexicon-based*.

## ***Fitur Ekstraksi***

Pada tahap ini setelah dilakukan pelabelan data, data hasil pelabelan selanjutnya akan dilakukan fitur ekstraksi. Fitur ekstraksi akan mengubah data berbentuk kata menjadi data berbentuk vektor. Tujuan dari fitur ekstraksi yaitu agar data bisa dibaca dan diolah oleh sistem komputer.

## ***Pembagian Data***

Pada tahap ini setelah dilakukan *pre-processing data*, data hasil *pre-processing* selanjutnya akan dilakukan pembagian data. Pembagian data *tweet* dilakukan dengan membagi data *tweet* hasil *pre-processing* yang sudah melalui tahap pelabelan menjadi dua, yaitu data *train* dan data *test*. Data *train* akan dipakai guna melatih model klasifikasi yang dibuat, sedangkan data *test* akan dipakai guna menguji model klasifikasi yang dibuat.

## ***Training Model***

Pada tahapan ini dibangun model LSTM (*Long Short-Term Memory*) dan model NBC (*Naïve Bayes Classifier*). Tahap *training* ini berguna sebagai bahan belajarnya model agar proses prediksi bisa dilakukan dengan baik menggunakan model tersebut. Sebelum dilakukan *training* model, dataset akan dirubah menjadi bentuk vektor. Kemudian dataset akan dipecah menjadi dua data, yang pertama data *testing* dan yang kedua data *training*. Banyaknya data *training* dan data *testing* yang akan diuji dengan beberapa variasi pembagian guna mencari nilai akurasi terbaik. Variasi yang dibuat yaitu data *training* dan data *testing* dengan perbandingan 90% : 10%, 80% : 20%, 70% : 30% dari total data. Setelah didapatkan data *training*, selanjutnya data *training* akan dipakai sebagai *input* dari tahap *training* model.

## Evaluasi Performa

Pada tahap ini digunakan *confusion matrix* untuk dilakukan evaluasi dari model. *Confusion matrix* menampilkan apa yang didapat dari pemahaman antar banyaknya data yang benar saat diprediksi dan banyaknya data yang tidak benar saat diprediksi dibandingkan dengan fakta yang didapatkan [6]. Perbandingan yang didapatkan akan dipakai untuk mendapatkan nilai *accuracy*, *precision*, *recall*, dan *f1 score* dari model yang sudah dibuat sebelumnya.

## Analisis Sentimen


Penelitian analisis sentimen dilakukan dengan penambahan teks yang bertujuan untuk analisis opini, pandangan, evaluasi, sentimen dari kebanyakan orang dalam bentuk teks untuk ditinjau dari suatu objek, sehingga dapat diketahui teks kalimat itu termasuk dalam kalimat yang bernilai positif ataupun negatif [7]. Pada tahap analisis sentimen peneliti melakukan analisis sentimen terhadap isu penundaan Pemilu 2024 dengan membandingkan metode *Long Short-Term Memory* dan *Naive Bayes Classifier*. Analisis sentimen dilakukan dengan mengklasifikasikan kalimat menjadi sentimen positif dan negatif. Adapun data yang digunakan untuk klasifikasi yaitu data yang berbeda dari data *training* dan data *testing*.

## HASIL DAN PEMBAHASAN

### Crawling Data

Tahap *crawling* data pada Twitter dilakukan menggunakan API dari Twitter dengan kata kunci “penundaan pemilu” dan “tunda pemilu. *Crawling* data dilakukan peneliti pada bulan Juni, September, dan Oktober 2022. Hasil dari *crawling* data ini terkumpul sebanyak 3.192 data mentah, kemudian data disimpan dalam format .csv, dan diberi kolom dengan nama “waktu, *user*, dan teks”. Sampel dari hasil *crawling* data yang dilakukan peneliti ditunjukkan seperti pada **TABEL 1**.

**TABEL 1.** Hasil *crawling* data

No	Waktu	User	Teks
1	18/09/2022 04:17	NephiLaxmus	@mpujayaprema @jokowi Tunda pemilu, 3 periode, cawapres, nambah tahun dll cuma modus2 opini. Pada intinya perpanjan... <a href="https://t.co/g6Fp1DFco8">https://t.co/g6Fp1DFco8</a>
2	17/09/2022 17:31	SarasaSariksa	Purnawirawan jendral, mantan presiden, orang paling buanyak duit nya, pasukan nya lengkapppp....  Berubah jadi Ayam... <a href="https://t.co/QAAiS4e2Et">https://t.co/QAAiS4e2Et</a>
3	21/09/2022 11:35	djokoAdv	Klo dicurigai tidak bakal adil, pemilu di tunda dulu..biar dilanjutkan oleh pak Jokowi hingga 2029....  <a href="https://t.co/DxlKYHjBWj">https://t.co/DxlKYHjBWj</a>

Pada **TABEL 1** terdapat kolom waktu yang menunjukkan kapan waktu *tweet* dibuat. Kolom *user* yang menunjukkan pengguna Twitter dari yang mengunggah *tweet*. Kolom teks yang menunjukkan isi dari *tweet* yang diunggah oleh pengguna Twitter.

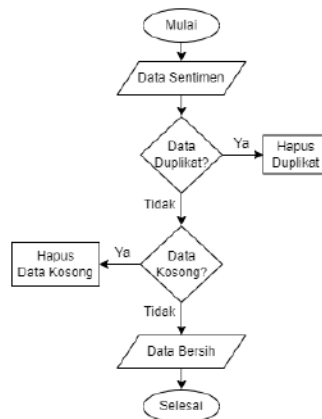
## Pre-processing Data

Pada tahap *pre-processing* dimulai dengan melakukan *cleaning data*, *case folding*, *tokenization*, normalisasi, *stopword removal*, dan *stemming*. *Pre-processing* dilakukan untuk mengubah data *tweet* dari hasil *crawling* menjadi data yang siap diolah. Hasil data yang sudah dilakukan *pre-processing* ditunjukkan seperti pada **TABEL 2**.

**TABEL 2.** Hasil *pre-processing*

No	Sebelum <i>Pre-processing</i>	Sesudah <i>Pre-processing</i>
1	@mpujayaprema @jokowi Tunda pemilu, 3 periode, cawapres, nambah tahun dll cuma modus2 opini. Pada intinya perpanjan... <a href="https://t.co/g6Fp1DFco8">https://t.co/g6Fp1DFco8</a>	tunda pilih umum periode cawapres nambah tahun dan lain modal dusta opini inti panjang
2	Purnawirawan jendral, mantan presiden, orang paling buanyak duit nya, pasukan nya lengkapppp....  Berubah jadi Ayam... <a href="https://t.co/QAAiS4e2Et">https://t.co/QAAiS4e2Et</a>	purnawirawan jendral mantan presiden uang pasu lengkap ubah ayam
3	Klo dicurigai tidak bakal adil, pemilu di tunda dulu biar dilanjutkan oleh pak Jokowi hingga 2029.... 🇮🇩 <a href="https://t.co/DxlKYHjBWj">https://t.co/DxlKYHjBWj</a>	curiga adil pemilu tunda biar lanjut jokowi

Pada **TABEL 2** terdapat kolom data *tweet* sebelum dilakukan *pre-processing* dan sesudah dilakukan *pre-processing*. Sebelum dilakukan *pre-processing*, data masih terdapat simbol, emoji, *mention*, tagar, spasi yang berlebih, huruf besar, kata yang tidak sesuai dengan KBBI, kata yang tidak mempunyai nilai sentimen, imbuhan kata. Setelah dilakukan *pre-processing*, data sudah lebih bersih. Kemudian dilakukan penghapusan data duplikat dan data kosong dilakukan agar tidak terdapat data yang berulang dan kosong, sehingga dapat mengurangi *noise* pada data. Tahapan penghapusan data duplikat dan data kosong yang dilakukan penulis ditunjukkan seperti pada **GAMBAR 2**.



**GAMBAR 2.** Tahapan hapus duplikat dan data kosong

Pada **GAMBAR 2** tahapan penghapusan data duplikat dan data kosong dimulai dengan dimasukkannya data sentimen, kemudian kondisi dari data sentimen apakah terdapat data duplikat, jika iya maka hapus duplikat, jika tidak maka masuk kondisi dari data sentimen apakah terdapat data kosong, jika iya maka hapus data kosong, jika tidak maka masuk data bersih sebagai keluaran, kemudian selesai. Hasil data setelah dilakukan proses *pre-processing* data didapatkan data bersih sebanyak 1.424 data.

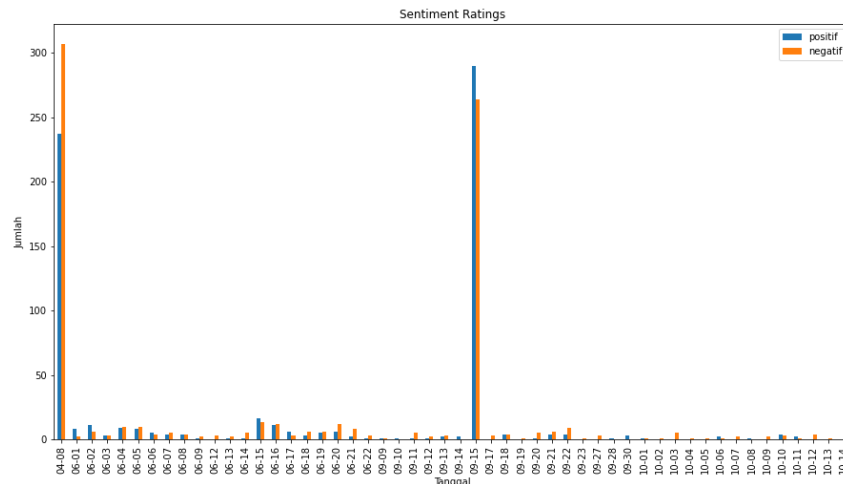
## Pelabelan Data

Pelabelan data diselesaikan dengan cara otomatis dengan menggunakan metode *Lexicon-based*. Pada metode ini digunakan kamus *Lexicon* yang diperoleh dari akun GitHub fajri91. Pada metode *Lexicon-based* setiap kata dalam kalimat diberikan nilai sentimen dan dijumlahkan, jika hasilnya lebih dari 0 maka kalimat itu memiliki sentimen positif, jika hasilnya kurang dari sama dengan 0 maka kalimat tersebut memiliki sentimen negatif. Hasil dari pelabelan data ditunjukkan seperti pada **TABEL 3**.

**TABEL 3.** Hasil pelabelan data

No	Stemming	Polarity Score	Sentimen
1	['tunda', 'pemilu', 'periode', 'cawapres', 'nambah', 'tahun', 'dan lain', 'modal dusta', 'opini', 'inti', 'panjang']	2	positif
2	['purnawirawan', 'jenderal', 'mantan', 'presiden', 'uang', 'pasu', 'lengkap', 'ubah', 'ayam']	-1	negatif
3	['curiga', 'adil', 'pemilu', 'tunda', 'biar', 'lanjut', 'jokowi']	-5	negatif

Pada **TABEL 3** baris nomor 1 didapat nilai sentimen sebesar 2 maka sentimen bernilai positif. Pada baris 2 didapat nilai sentimen sebesar -1 maka sentimen bernilai negatif. Pada baris 3 didapat nilai sentimen sebesar -5 maka sentimen bernilai negatif. Hasil setelah dilakukan pelabelan didapat *tweet* yang bernilai positif sebanyak 667 *tweet*, dan *tweet* yang bernilai negatif sebanyak 757 *tweet*. Adapun hasil persebaran data sentimen ditunjukkan seperti pada **GAMBAR 3**.



**GAMBAR 3.** Hasil persebaran data sentimen *Lexicon-Based*

Pada **GAMBAR 3** terdapat dua waktu yang mempunyai jumlah sentimen paling banyak, yaitu pada tanggal 08 April 2022 dengan jumlah sentimen positif lebih sedikit dari jumlah sentimen negatif, dan pada tanggal 15 September 2022 dengan jumlah sentimen positif lebih banyak dari jumlah sentimen negatif.

## Fitur Ekstraksi

Fitur Ekstraksi dilakukan dengan mengubah data berbentuk kata menjadi data berbentuk vektor. Tujuan dari fitur ekstraksi yaitu agar data bisa dibaca dan diolah oleh sistem komputer. Fitur ekstraksi yang digunakan pada metode klasifikasi *Long Short-Term Memory* (LSTM) yaitu dengan *word embedding*. Kemudian fitur ekstraksi yang digunakan pada metode klasifikasi *Naïve Bayes Classifier* (NBC) yaitu dengan *n-gram*. Hasil dari fitur ekstraksi ditunjukkan seperti pada **GAMBAR 4**.

```
array([[ 0,  0,  0, ..., 571, 572, 435],
       [ 0,  0,  0, ...,  2,  1,  6],
       [ 0,  0,  0, ..., 12,  2, 1245],
       ...,
       [ 0,  0,  0, ...,  9, 732, 820],
       [ 0,  0,  0, ..., 1194,  2,  1],
       [ 0,  0,  0, ...,  1,  61, 158]])
```

(a) Fitur ekstraksi *word embedding*

	abbas	abbas lapor	abdullah	absah	acara	aceh	aceh jaya	aceh timur
0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...
1420	0	0	0	0	0	0	0	0
1421	0	0	0	0	0	0	0	0
1422	0	0	0	0	0	0	0	0
1423	0	0	0	0	0	0	0	0
1424	0	0	0	0	0	0	0	0

(b) Fitur ekstraksi *n-gram*

GAMBAR 4. Hasil fitur ekstraksi

Pada GAMBAR 4 data *tweet* sudah diubah menjadi bentuk vektor atau angka, pada bagian label (a) merupakan hasil dari fitur ekstraksi *word embedding*, dan pada bagian label (b) merupakan hasil dari ekstraksi fitur *n-gram*.

### Pembagian Data

Pembagian data *tweet* dilakukan dengan memecah data *tweet* hasil *pre-processing* yang sudah melalui tahap pelabelan menjadi dua data, yaitu data *training* dan data *testing*. Data *training* digunakan untuk *training* model klasifikasi yang dibuat, sedangkan data *testing* digunakan untuk *testing* model klasifikasi yang dibuat. Data sebanyak 1.425 dari hasil *pre-processing* dan pelabelan akan dibagi menjadi tiga variasi pembagian data *training* dan data *testing*. Pembagian data *training* dan data *testing* ditunjukkan seperti pada TABEL 4.

TABEL 4. Hasil pembagian data

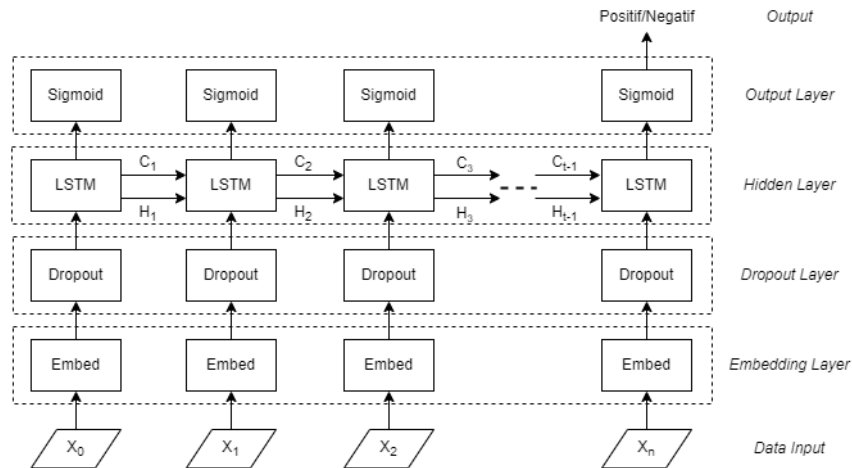
No	Metode	Pembagian Data	
		Training	Testing
1	<i>Long Short-Term Memory</i>	90%	10%
		80%	20%
		70%	30%
2	<i>Naïve Bayes Classifier</i>	90%	10%
		80%	20%
		70%	30%

Pada TABEL 4 metode *Long Short-Term Memory* dan *Naïve Bayes Classifier* perbandingan pembagian data *training* dan data *testing* masing-masing sebesar 90% : 10%, 80% : 20%, 70% : 30%.

### Training Model

*Training model* yang dilakukan pada penelitian ini digunakan dua metode klasifikasi. Metode pertama yaitu *Long Short-Term Memory* (LSTM) dan *Naïve Bayes Classifier* (NBC). Data *input* yang digunakan pada model klasifikasi dilakukan konversi dari data berbentuk kata menjadi data berbentuk vektor. Pada metode LSTM, terdapat beberapa tahapan yang dilakukan seperti yang ditunjukkan pada GAMBAR 5. Pada GAMBAR 5 tahapan LSTM dimulai dengan dimasukkannya data sentimen yang sudah berbentuk vektor atau angka. Kemudian menuju *layer input* atau *embedding layer*. Pada *layer* ini terdapat parameter *input dim* (ukuran maksimal dimensi kata) dengan nilai 10.000, *output dim* (dimensi *embedding*) dengan nilai 32, *input length* (batas maksimal panjang kalimat kata) dengan nilai 200. Kemudian menuju *dropout layer* dengan nilai 0.5 untuk mengurangi risiko *overfit*. Kemudian menuju *hidden layer*, pada *layer* ini akan didapatkan nilai yang akan dilupakan atau dibuang, nilai yang akan diperbarui menuju *cell state*, dan calon vektor baru yang akan dipakai pada *cell state*. Kemudian akan didapat nilai *cell state* yang terbaru,

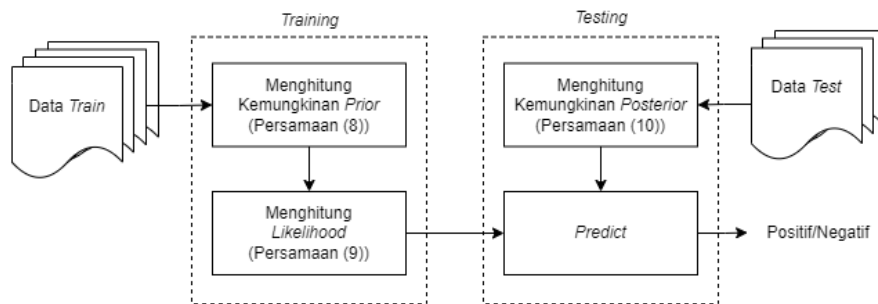
nilai *output* dan nilai dari *hidden state*. Kemudian menuju *output layer*, pada layer ini akan didapat *output* yaitu sentimen positif atau negatif.



**GAMBAR 5.** Tahapan LSTM

Untuk mengukur efektivitas model peneliti menggunakan *lost function binary\_crossentropy*. Kemudian setelah model dibangun, selanjutnya peneliti melakukan *training model* dengan *epoch* (banyaknya iterasi saat *training*) sebanyak 20, *batch\_size* (jumlah sampel setiap *epoch*) sebesar 64. Kemudian dilakukan *testing* menggunakan data *test* untuk didapatkan evaluasi model.

Pada metode NBC, terdapat beberapa tahapan yang dilakukan seperti yang ditunjukkan pada **GAMBAR 6**.



**GAMBAR 6.** Tahapan NBC

Pada **GAMBAR 6** tahapan NBC dimulai dengan dimasukkannya data sentimen yang sudah berbentuk vektor atau angka dari data *training*. Pada tahap *training* akan didapatkan nilai kemungkinan *prior* yaitu nilai dari banyaknya dokumen dari kelas tertentu. Kemudian nilai *likelihood* yaitu nilai kemungkinan kata yang muncul dalam kelas tertentu. Kemudian pada tahap *testing* akan didapatkan nilai kemungkinan *posterior*, sehingga didapat sentimen positif atau negatif dan evaluasi model.

### Evaluasi Performa

Evaluasi performa dilakukan dengan menggunakan *confusion matrix*. Pada *confusion matrix* terdapat nilai *accuracy*, nilai *precision*, nilai *recall*, dan nilai *f1-score*. Nilai dari evaluasi performa dari kedua metode yang digunakan penulis ditunjukkan seperti pada **TABEL 5**. Pada **TABEL 5** metode *Long Short-Term Memory* dengan variasi pembagian data 90% : 10% memperoleh nilai *accuracy* senilai 92%, *precision* pada kelas negatif senilai 92% dan kelas positif senilai 92%, *Recall* pada kelas negatif senilai 92% dan kelas positif senilai 92%, *F1-Score* pada kelas negatif senilai 92% dan kelas positif senilai 92%. Variasi pembagian data 80% : 20% memperoleh nilai *accuracy*



senilai 87%, *precision* pada kelas negatif senilai 89% dan kelas positif senilai 85%, *Recall* pada kelas negatif senilai 84% dan kelas positif senilai 90%, *F1-Score* pada kelas negatif senilai 86% dan kelas positif senilai 87%. Variasi pembagian data 70% : 10% memperoleh nilai *accuracy* senilai 83%, *precision* pada kelas negatif senilai 85% dan kelas positif senilai 81%, *Recall* pada kelas negatif senilai 81% dan kelas positif senilai 85%, *F1-Score* pada kelas negatif senilai 83% dan kelas positif senilai 83%.

Pada metode *Naïve Bayes Classifier* dengan variasi pembagian data 90% : 10% memperoleh nilai *accuracy* senilai 80%, *precision* pada kelas negatif senilai 83% dan kelas positif senilai 77%, *Recall* pada kelas negatif senilai 79% dan kelas positif senilai 82%, *F1-Score* pada kelas negatif senilai 81% dan kelas positif senilai 80%. Variasi pembagian data 80% : 20% memperoleh nilai *accuracy* senilai 76%, *precision* pada kelas negatif senilai 77% dan kelas positif senilai 74%, *Recall* pada kelas negatif senilai 79% dan kelas positif senilai 72%, *F1-Score* pada kelas negatif senilai 78% dan kelas positif senilai 73%. Variasi pembagian data 70% : 30% memperoleh nilai *accuracy* senilai 72%, *precision* pada kelas negatif senilai 74% dan kelas positif senilai 70%, *Recall* pada kelas negatif senilai 75% dan kelas positif senilai 69%, *F1-Score* pada kelas negatif senilai 75% dan kelas positif senilai 69%.

TABEL 5. Hasil evaluasi performa

Metode	Data		Accuracy	Precision		Recall		F1-Score	
	Train	Test		0	1	0	1	0	1
	Long Short-Term Memory	90%		10%	0.92	0.92	0.92	0.92	0.92
80%		20%	0.87	0.89	0.85	0.84	0.90	0.86	0.87
70%		30%	0.83	0.85	0.81	0.81	0.85	0.83	0.83
Naïve Bayes Classifier	90%	10%	0.80	0.83	0.77	0.79	0.82	0.81	0.80
	80%	20%	0.76	0.77	0.74	0.79	0.72	0.78	0.73
	70%	30%	0.72	0.74	0.70	0.75	0.69	0.75	0.69

### Analisis Sentimen

Analisis sentimen dilaksanakan dengan mengklasifikasikan kalimat menjadi sentimen positif dan negatif. Klasifikasi dilakukan dengan memakai model *Long Short-Term Memory* dan *Naïve Bayes Classifier* yang sudah dibuat dengan nilai akurasi paling tinggi dari masing-masing model. Model *Long Short-Term Memory* menggunakan variasi pembagian data 90% : 10% dengan akurasi 92%, dan model *Naïve Bayes Classifier* menggunakan variasi pembagian data 90% : 10% dengan akurasi 80%. Adapun data yang digunakan untuk klasifikasi yaitu data yang berbeda dari data *training* dan data *testing*, data klasifikasi ini diambil pada tanggal 15 dan 23 Desember 2022. Data baru digunakan dalam klasifikasi ditunjukkan seperti pada **GAMBAR 7**.

	waktu	tweets
0	2022-12-23 03:01:06	picu konflik sara sengaja lolos cipta kondisi ...
1	2022-12-23 02:57:23	jokowi milu sportif demokrat upaya tunda milu
2	2022-12-23 02:45:22	tahap tunda milu
3	2022-12-23 02:45:17	hubung gara ketua kpu cabul milu tunda
4	2022-12-23 02:43:29	tunda
...	...	...
645	2022-12-13 23:19:42	tuju cerita tunda milu wacana periode
646	2022-12-13 22:52:45	negara nenek moyang kudeta konstitusi undang b...
647	2022-12-13 22:23:28	cebong ngotot tunda milu periode iya iya kasih...
648	2022-12-13 22:04:54	harus milu tundatetapi tiadakansoal dpn matara...
649	2022-12-13 22:03:34	memble ikn tunda milu periode

650 rows × 2 columns

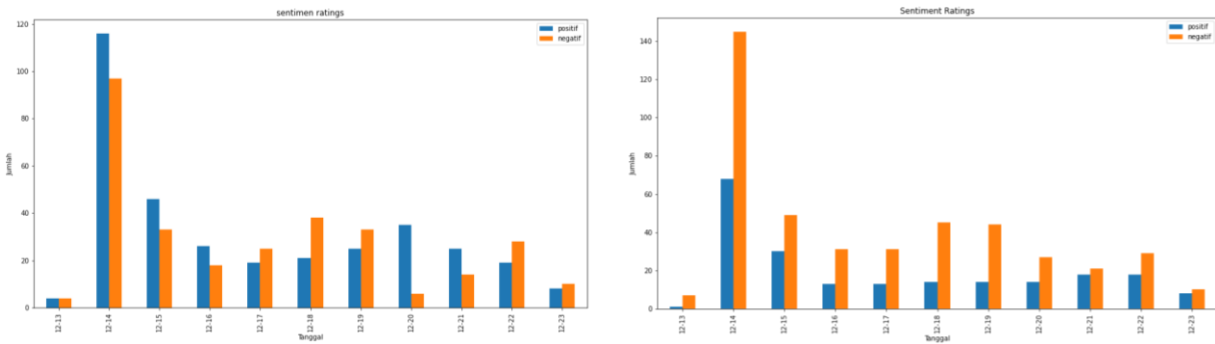
GAMBAR 7. Data baru

Pada **GAMBAR 7** terdapat 650 data yang sudah bersih dan terdapat kolom waktu serta kolom *tweets*. Pada kolom waktu merupakan waktu *tweet* dibuat dan pada kolom *tweets* merupakan isi dari unggahan *tweet*. Hasil yang di dapat dari klasifikasi sentimen dari model *Long Short-Term Memory* dan *Naïve Bayes Classifier* ditunjukkan seperti pada **GAMBAR 8**.



(a) Hasil klasifikasi sentimen LSTM (b) Hasil klasifikasi sentimen NBC  
**GAMBAR 8.** Hasil klasifikasi sentimen

Pada **GAMBAR 8** bagian label (a) merupakan hasil klasifikasi sentimen dari model *Long Short-Term Memory* dan didapatkan klasifikasi positif sebesar 52.9% dan klasifikasi negatif sebesar 47.1%. Sedangkan bagian label (b) merupakan hasil klasifikasi sentimen dari model *Naïve Bayes Classifier* dan didapatkan klasifikasi positif sebesar 32.5% dan klasifikasi negatif sebesar 67.5%. Adapun hasil persebaran data sentimen ditunjukkan seperti pada **GAMBAR 9**.



(a) Hasil persebaran data sentimen LSTM (b) Hasil persebaran data sentimen NBC  
**GAMBAR 9.** Hasil persebaran data sentimen

Pada **GAMBAR 9** bagian label (a) merupakan hasil persebaran data sentimen dari model *Long Short-Term Memory* dan terdapat data sentimen paling sedikit yaitu pada tanggal 13 Desember 2022 dengan jumlah sentimen positif sama dengan jumlah sentimen negatif, dan data sentimen paling banyak yaitu pada tanggal 14 Desember 2022 dengan jumlah sentimen positif lebih banyak dari jumlah sentimen negatif. Sedangkan bagian label (b) merupakan hasil persebaran data sentimen dari model *Naïve Bayes Classifier* dan terdapat data sentimen paling sedikit yaitu pada tanggal 13 Desember 2022 dengan jumlah sentimen positif lebih sedikit dari jumlah sentimen negatif, dan data sentimen paling banyak yaitu pada tanggal 14 Desember 2022 dengan jumlah sentimen positif lebih sedikit dari jumlah sentimen negatif.

## KESIMPULAN

Analisis sentimen pada Twitter terhadap isu penundaan Pemilu 2024 dengan metode *Long Short-Term Memory* didapat sentimen positif sebesar 52.9% dan sentimen negatif sebesar 47.1%, sedangkan dengan metode *Naïve Bayes Classifier* didapat sentimen positif sebesar 67.5% dan sentimen negatif sebesar 32.5%. Akurasi yang didapat dari

metode *Long Short-Term Memory* sebesar 92%, sedangkan akurasi yang didapat dari metode *Naïve Bayes Classifier* sebesar 80%.

## REFERENSI

- [1] G. N. Aulia dan E. Patriya, "Implementasi Lexicon Based Dan Naive Bayes Pada Analisis Sentimen Pengguna Twitter Topik Pemilihan Presiden 2019," *J. Ilm. Inform. Komput.*, vol. 24, no. 2, hal. 140–153, 2019, doi: 10.35760/ik.2019.v24i2.2369.
- [2] Y. Nooryuda Prasetya dan D. Winarso, "Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19," *J. Fasilkom*, vol. 11, no. 2, hal. 97–103, 2021.
- [3] C. K. N. Papatungan dan A. Jacobus, "Sentiment Analysis of Social Media Users Using Long-Short Term Memory Method," *J. Tek. Elektro dan Komput.*, vol. 10, no. 2, hal. 99–106, 2021.
- [4] A. Kusuma dan A. Nugroho, "Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes," *J. Ilm. Teknol. Inf. Asia*, vol. 15, no. 2, hal. 137, 2021, doi: 10.32815/jitika.v15i2.557.
- [5] R. I. N. M. Wijaya, Tamora Nonia, "Analisis Sentimen Opini Publik Tentang Undang- Undang Cipta Kerja Pada Twitter," *Jambura J. Electr. Electron. Eng.*, vol. 3, hal. 78–83, 2021.
- [6] Merinda Lestandy, Abdurrahim Abdurrahim, dan Lailis Syafa'ah, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, hal. 802–808, 2021, doi: 10.29207/resti.v5i4.3308.
- [7] Y. Romadhoni *dkk.*, "Analisis Sentimen Terhadap PERMENDIKBUD No . 30 pada Media Sosial Twitter Menggunakan Metode Naive Bayes dan LSTM," vol. 7, no. 2, hal. 118–124, 2022.
- [8] Permana, Rakhmad Hidayatulloh. (2022, April 07). "Timeline Wacana Penundaan Pemilu: Digaungkan Para Menteri, Ditutup Jokowi". Retrieved from <https://news.detik.com/berita/d-6020931/timeline-wacana-penundaan-pemilu-digaungkan-para-menteri-ditutup-jokowi>.
- [9] Rizaty, Monavia Ayu. (2022). "Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022". Retrieved from <https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022>