

Analisis Sentimen Menggunakan Metode Neural Network(NN) dan Long Short Term Memory (LSTM)

Team members:

1. Arief Budiman
2. Muhidin Jaariyah
3. Rommy



C O N T E N T S

01.

Pendahuluan

02.

Metode Penelitian

03.

Kesimpulan

PENDAHULUAN





Latar belakang

Analisis sentimen merupakan proses penting dalam menganalisis opini, perasaan, dan sikap yang terkandung dalam teks. Dalam era digital saat ini, jumlah data teks yang dihasilkan melalui media sosial, ulasan pelanggan, dan sumber lainnya semakin meningkat. Untuk mengatasi kompleksitas dan volume besar data tersebut, pendekatan yang digunakan adalah jaringan saraf tiruan (neural network) dan Long short term memory (LSTM) telah terbukti berhasil dalam analisis sentimen.



Rumusan masalah

Dalam konteks analisis sentimen dengan menggunakan NN dan LSTM terdapat beberapa rumusan masalah yang dapat diajukan, antara lain:

1. Bagaimana cara menerapkan analisis sentimen menggunakan NN dan LSTM untuk mengklasifikasikan teks ke dalam tiga kategori sentimen: positif, negatif, dan netral?
2. Seberapa akurat dan efektifkah metode analisis sentimen berbasis NN dan LSTM dalam mengklasifikasikan sentimen pada teks?



Tujuan penelitian

1. Menerapkan metode analisis sentimen berbasis NN dan LSTM untuk mengklasifikasikan teks ke dalam tiga kategori sentimen: positif, negatif, dan netral.
2. Mengevaluasi akurasi dan efektivitas metode analisis sentimen berbasis NN dan LSTM dalam memprediksi sentimen pada teks, dengan membandingkannya dengan metode tradisional.
3. Memberikan pemahaman yang lebih baik tentang penggunaan NN dan LSTM dalam analisis sentimen dan potensinya untuk mendukung pengambilan keputusan berbasis data.

Metode Penelitian



All Data

	Text	Sentimen
0	warung ini dimiliki oleh pengusaha pabrik tahu...	positive
1	mohon ulama lurus dan k212 mmbri hujjah partai...	neutral
2	lokasi strategis di jalan sumatera bandung . t...	positive
3	betapa bahagia nya diri ini saat unboxing pake...	positive
4	duh . jadi mahasiswa jangan sombong dong . kas...	negative
...
10995	tidak kecewa	positive
10996	enak rasa masakan nya apalagi kepiting yang me...	positive
10997	hormati partai-partai yang telah berkoalisi	neutral
10998	pagi pagi di tol pasteur sudah macet parah , b...	negative
10999	meskipun sering belanja ke yogya di riau junct...	positive
11000 rows × 2 columns		

Cleansing Data

	Text	Sentimen	text_clean	text_filter
0	warung ini dimiliki oleh pengusaha pabrik tahu...	positive	warung ini dimiliki oleh pengusaha pabrik tahu...	warung dimiliki pengusaha pabrik puluhan terke...
1	mohon ulama lurus dan k212 mmbri hujjah partai...	neutral	mohon ulama lurus dan k212 mmbri hujjah partai...	mohon ulama lurus k212 mmbri hujjah partai diw...
2	lokasi strategis di jalan sumatera bandung . t...	positive	lokasi strategis di jalan sumatera bandung t...	lokasi strategis jalan sumatera bandung nya ny...
3	betapa bahagia nya diri ini saat unboxing pake...	positive	betapa bahagia nya diri ini saat unboxing pake...	betapa bahagia nya unboxing paket barang nya b...
4	duh . jadi mahasiswa jangan sombong dong . kas...	negative	duh jadi mahasiswa jangan sombong dong kas...	duh mahasiswa sombong kasih kartu kuning belaj...
...
10995	tidak kecewa	positive	tidak kecewa	kecewa
10996	enak rasa masakan nya apalagi kepiting yang me...	positive	enak rasa masakan nya apalagi kepiting yang me...	enak masakan nya kepiting menyenangkan memilih...
10997	hormati partai-partai yang telah berkoalisi	neutral	hormati partai partai yang telah berkoalisi	hormati partai partai berkoalisi
10998	pagi pagi di tol pasteur sudah macet parah , b...	negative	pagi pagi di tol pasteur sudah macet parah b...	pagi pagi tol pasteur macet parah bikin jengkel
10999	meskipun sering belanja ke yogya di riau junct...	positive	meskipun sering belanja ke yogya di riau junct...	belanja yogya riau junction kali lihat foodlif...
11000 rows × 4 columns				

Feature Extraction

```
# df_vektor = df.text_filter.tolist()  
df_vektor = df.text_filter.tolist()
```

```
count_vect = CountVectorizer()
```

```
# melakukan fitting dan transformasi pada dokumen  
count_vect.fit(df_vektor)
```

```
# melihat hasil representasi bag of words  
X = count_vect.fit_transform(df_vektor)
```

Evaluasi Data Training

	precision	recall	f1-score	support
negative	0.91	0.96	0.94	2775
neutral	0.95	0.89	0.92	915
positive	0.97	0.96	0.97	5110
accuracy			0.95	8800
macro avg	0.95	0.94	0.94	8800
weighted avg	0.95	0.95	0.95	8800

Evaluasi Data Testing

	precision	recall	f1-score	support
negative	0.79	0.85	0.82	661
neutral	0.82	0.70	0.76	233
positive	0.92	0.91	0.91	1306
accuracy			0.87	2200
macro avg	0.84	0.82	0.83	2200
weighted avg	0.87	0.87	0.87	2200

All Data

	text	label
0	warung ini dimiliki oleh pengusaha pabrik tahu...	positive
1	mohon ulama lurus dan k212 mmbri hujjah partai...	neutral
2	lokasi strategis di jalan sumatera bandung . t...	positive
3	betapa bahagia nya diri ini saat unboxing pake...	positive
4	duh . jadi mahasiswa jangan sombong dong . kas...	negative

Cleansing Data

	text	negative	neutral	positive	text_new
0	warung ini dimiliki oleh pengusaha pabrik tahu...	0	0	1	warung ini dimiliki oleh pengusaha pabrik tahu...
1	mohon ulama lurus dan k212 mmbri hujjah partai...	0	1	0	mohon ulama lurus dan k212 mmbri hujjah partai...
2	lokasi strategis di jalan sumatera bandung . t...	0	0	1	lokasi strategis di jalan sumatera bandung . t...
3	betapa bahagia nya diri ini saat unboxing pake...	0	0	1	betapa bahagia nya diri ini saat unboxing pake...
4	duh . jadi mahasiswa jangan sombong dong . kas...	1	0	0	duh . jadi mahasiswa jangan sombong dong . kas...

Feature Extraction

```
x = df_baru['text_new'].values  
y = df_baru[['negative', 'neutral', 'positive']].values
```

```
220/220 [=====] - 2s 8ms/step - loss: 0.2792 - categorical_accuracy: 0.9043  
55/55 [=====] - 0s 8ms/step - loss: 0.3308 - categorical_accuracy: 0.8761  
69/69 [=====] - 1s 8ms/step - loss: 0.3481 - categorical_accuracy: 0.8709
```

```
Train: 0.904, val: 0.876, test: 0.871,
```

```
tokenizer = Tokenizer(num_words=5000, oov_token='x')  
tokenizer.fit_on_texts(x)  
sekuens_x = tokenizer.texts_to_sequences(x)  
padded_x = pad_sequences(sekuens_x)
```

Evaluasi Data Training

```
220/220 [=====] - 2s 8ms/step - loss: 0.2792 - categorical_accuracy: 0.9043
```

```
55/55 [=====] - 0s 8ms/step - loss: 0.3308 - categorical_accuracy: 0.8761
```

```
69/69 [=====] - 1s 8ms/step - loss: 0.3481 - categorical_accuracy: 0.8709
```

```
Train: 0.904, val: 0.876, test: 0.871,
```


Kesimpulan



KESIMPULAN

Untuk Model NN

Akurasi model NN untuk data training adalah 95% dan untuk data testing sebesar 87%. Ini menunjukkan bahwa model memiliki tingkat keakuratan yang tinggi dalam melakukan klasifikasi.

Untuk Model LSTM

Akurasi model LSTM untuk data training adalah 90,4% dan untuk data testing sebesar 87,1%. Ini menunjukkan bahwa model memiliki tingkat keakuratan yang tinggi dalam melakukan klasifikasi.

THANKS

