

...

Sentimen Analysis

M Ridha - Novitasari - Dwiky A

Pengertian

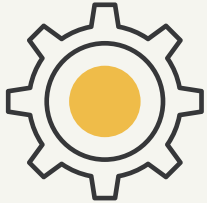


- Sentiment analysis adalah proses mengidentifikasi perasaan dan pendapat dalam teks.
- Sentiment analysis menggunakan NLP untuk mengidentifikasi sentimen dan memberikan interpretasi hasilnya.

LATAR BELAKANG



Penggunaan media sosial di Indonesia yang tinggi menyediakan banyak data untuk sentiment analysis.



Teknologi natural language processing (NLP) telah berkembang sehingga memudahkan sentiment analysis di media sosial.



Sentiment analysis di media sosial di Indonesia populer karena dapat memberikan wawasan tentang tanggapan pelanggan dan opini publik.

”

MANFAAT

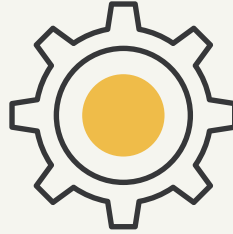
Sentiment analysis di media sosial dapat membantu mengetahui tanggapan pelanggan terhadap layanan, memahami opini publik, dan mengidentifikasi kekuatan dan kelemahan layanan.



AGENDA



DATA



MODELLING



**KESIMPULAN
&
REKOMENDASI**



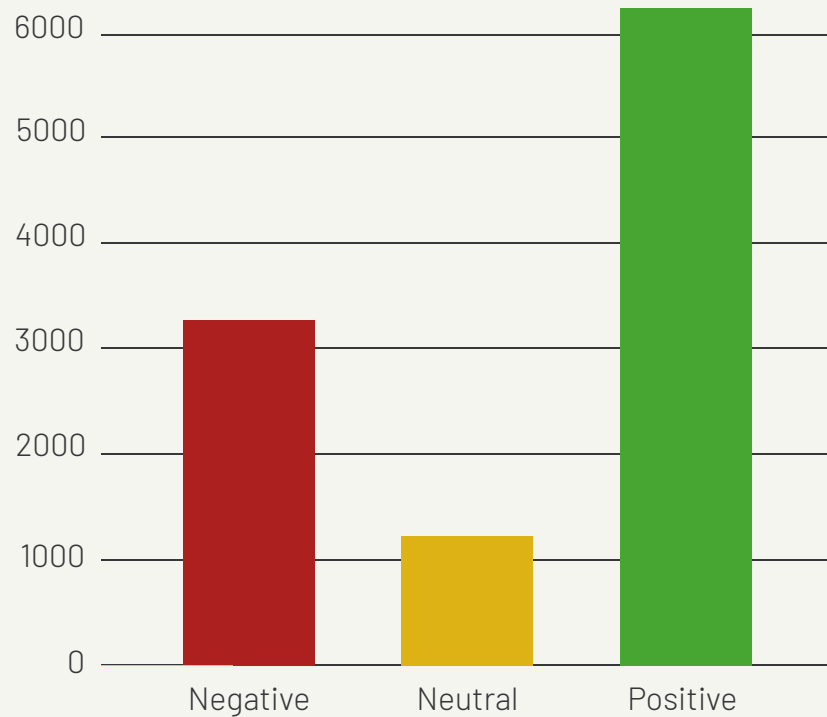
DATA





Original

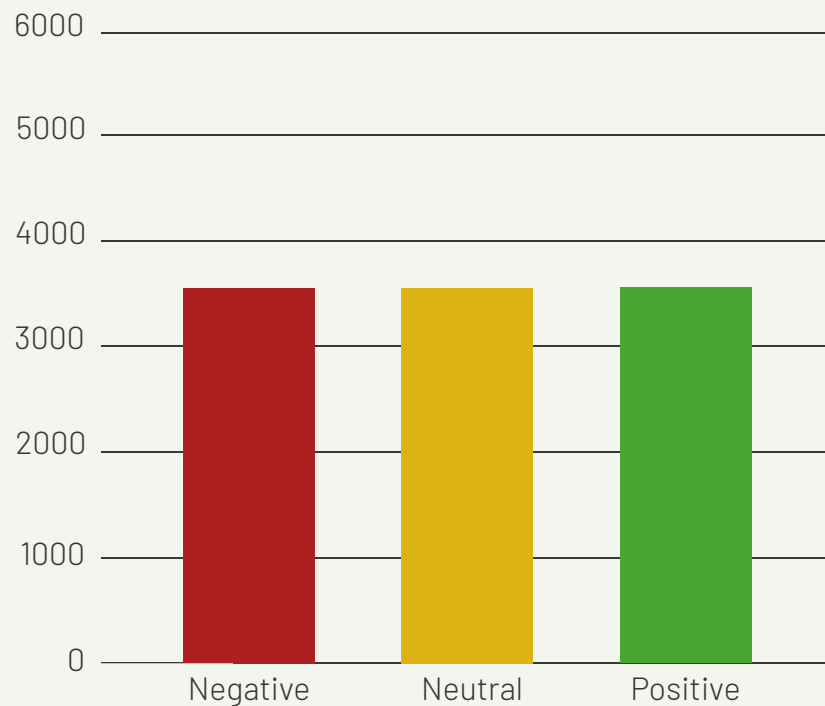
Data





Balancing

Data





Average Character & Word per sentiment

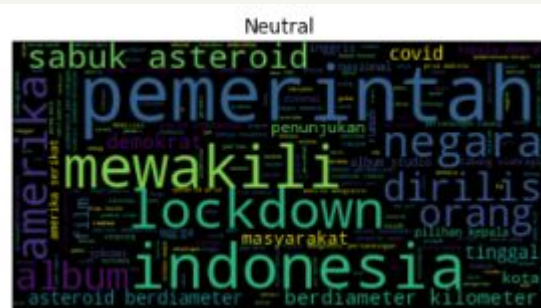
Original	Negative	Neutral	Positive
Avg. Word	23.755821	14.733449	41.101153
Avg. Character	113.582654	75.181185	194.832294
Balance	Negative	Neutral	Positive
Avg. Word	23.765000	17.490000	41.138824
Avg. Character	113.638529	101.875000	195.282059

Pengguna rata-rata lebih banyak menuliskan kata-kata dalam reviews positive dibandingkan ketika menulis reviews negative dan netral

Word Cloud



Negative



Neutral



Positive

CLEANSING TEXT

Normalize

```
# remove kata alay
def normalize_alay(text):
    alay_dict_map = dict(zip(dfAlay['original'], dfAlay['replacement']))
    return ' '.join([alay_dict_map[word] if word in alay_dict_map else word for word in text.split(' ')])
```

Punctuation

```
# remove punctuation
def remove_punctuation(text):
    return re.sub(r'^\w\s', ' ', text)
```

Stopword

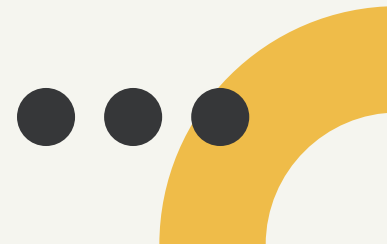
```
# remove stopwords
def remove_stopword(text):
    stopword_list = set(stopwords.words('indonesian'))
    return ' '.join([word for word in text.split(' ') if word not in stopword_list])
```

Lowercase

```
dfText['text'] = dfText['text'].str.lower()
```



DATA PREPROCESSING



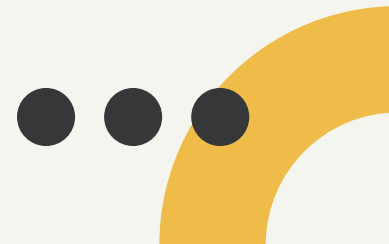
Before

Index	Text	Label
910	saya menyesal beli ipx . bahkan sampai harus menghitung . ternyata biasa saja...	Negative
911	salah satu sudut kota bandung terdapat beer house bernama q meal di daerah ...	Positive
912	menu nya beragam . waktu itu saya mencoba beli pizza nya yang besar . ukuran ...	Positive
913	sudah lama dibaned , kalau pakai telkom dan telkomsel , sedy ya : ... melete sih ...	Negative
914	ridwan kamil klaim silaturahmi dengan pdip masih tahap awal ...	Neutral

After

Index	Text	Label
910	menyesal beli ipx menghitung telepon genggam nya...	Negative
911	salah sudut kota bandung bir house bernama meal daerah ciumbuleuit nyaman...	Positive
912	menu nya beragam mencoba beli pizza nya ukuran pizza nya tipis pizza nya ...	Positive
913	dibaned pakai telkom telkomsel sedy ya melete sih 2012...	Negative
914	ridwan kamil klaim silaturahmi partai demokrasi indonesia perjuangan tahap...	Neutral

FEATURE EXTRACTION



TF-IDF

Metode yang mengekspresikan bobot kata dalam dokumen teks dengan menggunakan rumus yang mengukur frekuensi kemunculan kata dalam dokumen dan di seluruh dokumen yang tersedia.

Train Test Split

Teknik yang membagi dataset menjadi data training dan data testing untuk membangun dan mengevaluasi model pembelajaran mesin.

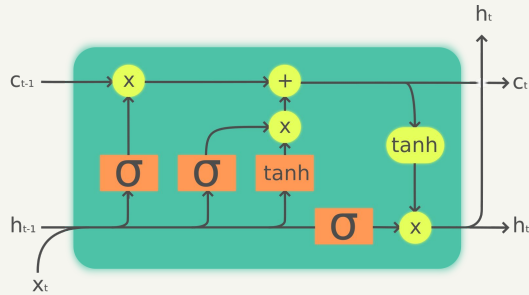
Tokenization and Change Label to Categorical

Tokenization memecah teks menjadi token kecil, sementara Change Label to Categorical merubah label teks menjadi kategori untuk mempersiapkan data teks untuk model pembelajaran mesin.



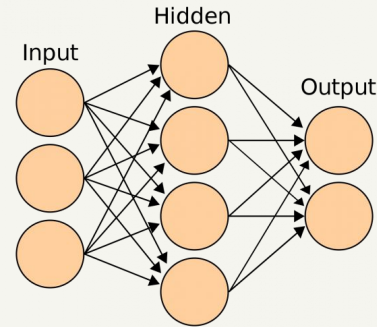
MODELLING

MODELLING



LSTM

LSTM (Long Short-Term Memory) adalah jenis jaringan saraf tiruan yang dapat mengingat informasi jangka panjang dan mengabaikan informasi yang tidak relevan.



ANN

ANN (Artificial Neural Network) adalah jaringan saraf tiruan yang terdiri dari neuron-neuron yang terhubung satu sama lain dan dapat mengatur cara menggunakan informasi untuk memprediksi output atau mengambil keputusan.

LSTM

	precision	recall	f1-score	support
Neutral	0.73	0.61	0.66	292
Negative	0.85	0.91	0.88	1594
Positive	0.76	0.71	0.73	864
accuracy			0.81	2750
macro avg	0.78	0.74	0.76	2750
weighted avg	0.81	0.81	0.81	2750

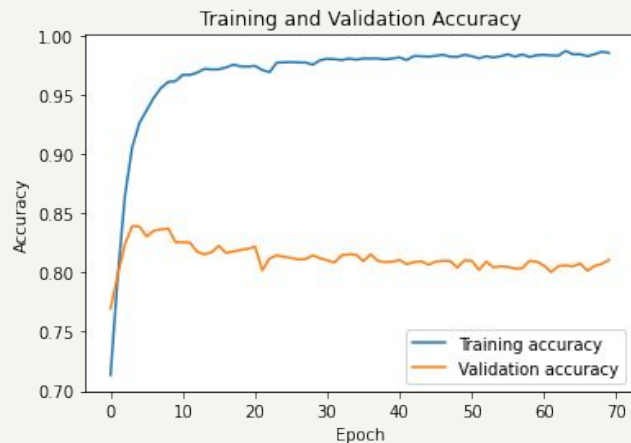
V/S

	precision	recall	f1-score	support
Neutral	0.92	0.89	0.90	881
Negative	0.79	0.85	0.82	839
Positive	0.79	0.76	0.77	830
accuracy			0.83	2550
macro avg	0.83	0.83	0.83	2550
weighted avg	0.83	0.83	0.83	2550

Original

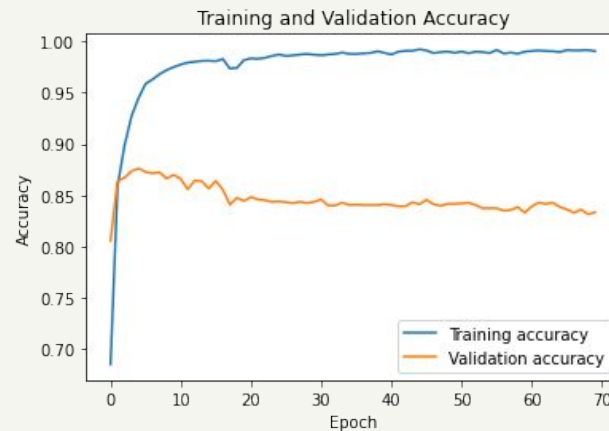
Balance

LSTM



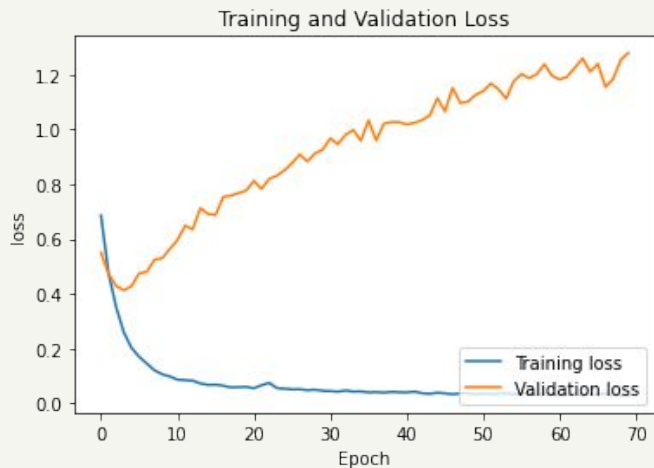
Original

V/S



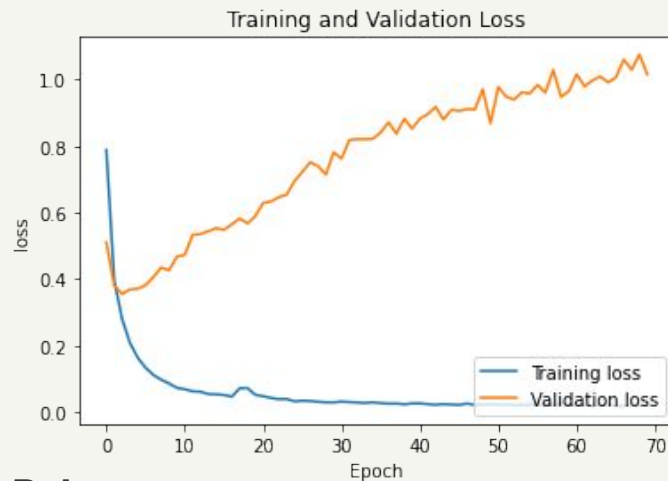
Balance

LSTM



Original

V/S



Balance



Confussion Matrix

Negative Sentiment

- TP = 0.86
- FP = $0.0091 + 0.09 = 0.0991$
- FN = $0.11 + 0.03 = 0.14$
- TN = $0.95 + 0.078 + 0.83 + 0.044 = 2.298$

Neutral Sentiment

- TP = 0.95
- FP = $0.03 + 0.078 = 0.108$
- FN = $0.0091 + 0.044 = 0.0531$
- TN = $0.86 + 0.11 + 0.09 + 0.83 = 1.89$

Positive Sentiment

- TP = 0.83
- FP = $0.044 + 0.11 = 0.154$
- FN = $0.078 + 0.09 = 0.168$
- TN = $0.95 + 0.0091 + 0.03 + 0.86 = 1.7591$



HASIL PREDIKSI LSTM

predict sentiment analysis in Bahasa Indonesia

Which machine learning model to use?

LSTM

Enter a sentence/word in Bahasa Indonesia

warung ini makannya enak

Upload CSV



Drag and drop file here

Limit 200MB per file • CSV, TXT, TSV

Browse files

your result:

```
{
  "result_sentiment": "positive"
  "text": "warung ini makannya enak"
}
```

ANN

	precision	recall	f1-score	support
0	0.74	0.58	0.65	292
1	0.88	0.88	0.88	1594
2	0.76	0.76	0.76	864
micro avg	0.83	0.81	0.82	2750
macro avg	0.79	0.74	0.76	2750
weighted avg	0.83	0.81	0.82	2750
samples avg	0.81	0.81	0.81	2750

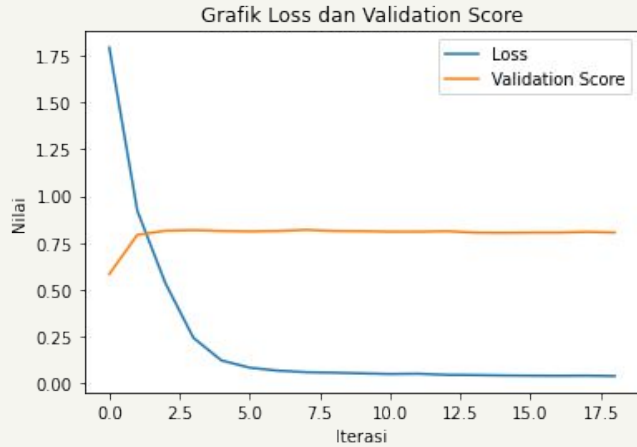
Original

V/S

	precision	recall	f1-score	support
0	0.94	0.91	0.93	881
1	0.86	0.86	0.86	839
2	0.84	0.80	0.82	830
micro avg	0.88	0.86	0.87	2550
macro avg	0.88	0.86	0.87	2550
weighted avg	0.88	0.86	0.87	2550
samples avg	0.85	0.86	0.86	2550

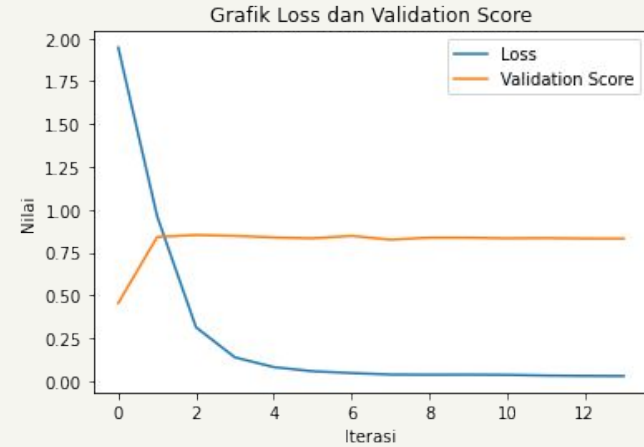
Balance

ANN



Original

V/S



Balance

HASIL PREDIKSI ANN

predict sentiment analysis in Bahasa Indonesia

Which machine learning model to use?

ANN

Enter a sentence/word in Bahasa Indonesia

Warung ini kotor

Upload CSV



Drag and drop file here

Limit 200MB per file • CSV, TXT, TSV

Browse files

your result:

```
{
  "result_sentiment": "negative"
  "text": "warung ini kotor"
}
```



KESIMPULAN & REKOMENDASI

KESIMPULAN



- Metode ANN (Artificial Neural Network) dan LSTM (Long Short-Term Memory) masing-masing memberikan hasil yang cukup baik dalam klasifikasi teks. Namun, pada kasus yang sama, ANN memberikan hasil yang lebih baik dari pada LSTM.
- Balancing data memiliki pengaruh yang signifikan terhadap hasil klasifikasi. Pada kasus yang sama, data yang telah di-balancing memberikan hasil yang lebih baik dari pada data yang belum di-balancing.
- Precision, recall, dan f1-score pada setiap kelas cukup tinggi, menandakan bahwa metode yang digunakan mampu mengelompokkan teks ke dalam kelas yang tepat dengan tingkat keakuratan yang tinggi.

REKOMENDASI



Untuk meningkatkan akurasi klasifikasi teks, disarankan untuk melakukan balancing terhadap data yang digunakan. Hal ini dapat dilakukan dengan menambah data untuk kelas yang minoritas atau mengurangi data kelas yang mayoritas.



Optimalkan hyperparameter yang digunakan agar hasil klasifikasi lebih baik.



Selain itu, dapat dipertimbangkan untuk menggunakan metode lain selain ANN dan LSTM, terutama jika hasil yang diinginkan belum tercapai dengan kedua metode tersebut.



Thank you!

