

Analisis Sentimen Berbasis Model Lexicon Berdasarkan Metode VADER

**Handoko Wisnu Murti
19/444054/TK/49250**

<https://screenrant.com/star-wars-darth-vader-deserves-movie-doesnt/>



Apa itu Analisis Sentimen?

Analisis sentimen adalah metode menganalisis kata-kata suatu teks untuk ditinjau tingkatan emosinya (Positif, negatif, atau netral)

Sentiment Analysis



Positif

Pelayanannya bagus dan
harganya terjangkau



Netral

Kemarin saya memesan
2 produk di aplikasi ini



Negatif

Barang yang sampai kondisinya
rusak. Pengiriman nya lama

**Analisis sentimen digunakan untuk memahami emosi atas
suatu opini**

Lebih Lanjut Mengenai Analisis Sentimen

04.

- Analisis sentimen mengacu pada suatu metode yang mengekstraksi subjektivitas dan polaritas suatu teks
- Suatu teks dapat dikatakan **subjektif** jika teks tersebut mengandung informasi non-faktual seperti pendapat pribadi, prediksi, dan justifikasi.
- Teks dapat dikatakan **objektif** jika teks tersebut mengandung informasi fakta ketimbang opini pribadi
- **Polaritas** suatu teks direpresentasikan dengan nilai yang memiliki rentang -1 hingga 1. -1 berarti memiliki emosi negatif, sedangkan 1 memiliki emosi positif.

Sentimen Positif: Polaritas > 0

Sentimen Netral: Polaritas $= 0$

Sentimen Negatif: Polaritas < 0

Untuk Tujuan apa Analisis Sentimen Biasa digunakan?

Untuk Mengetahui Kepuasan Pengguna

Kepuasan pengguna adalah hal utama yang perlu diperhatikan. Karena baik atau buruknya suatu opini publik akan mengakibatkan dampak yang signifikan pada suatu produk atau layanan

Dengan mengetahui kepuasan pengguna atas suatu layanan atau produk, suatu perusahaan/instansi bisa menentukan langkah tindak lanjut mengenai pengembangan atas layanan ataupun produk yang dimiliki

Bagaimana cara menentukan emosi suatu teks?

06.

Terdapat dua jenis metode untuk menentukan emosi suatu teks

Metode Leksikal

- Sentimen suatu kata ditentukan berdasarkan aturan penilaian yang sudah ditentukan.
- Metode ini telah memiliki kumpulan dictionary yang berisi berbagai kata disertai dengan nilai polaritas sentimennya
- Terdapat berbagai model berbasis metode leksikal yang memiliki aturan penilaiannya masing-masing. Contoh: TextBlob dan VADER

Metode Klasifikasi Machine Learning

- Sentimen suatu kata diklasifikasikan dengan menggunakan model machine learning
- Dataset training yang digunakan telah memiliki label sentimennya
- Penentuan sentimen dilakukan dengan memprediksi suatu teks berdasarkan model machine learning yang telah dibangun
- Dapat menggunakan model machine learning tipe klasifikasi

Metode Leksikal

Kata	Nilai Valensi
Great	3.1
Care	2.2
Perfect	2.7
Perfectionist	1.5
Impolite	-1.6
Bad	-2.5

- Setiap kata memiliki nilainya masing-masing.
- Setiap model memiliki penilaian yang berbeda-beda
- Tabel tersebut merujuk pada penilaian kata berdasarkan model **VADER**
- Nilai valensi pada tabel menunjukkan skor sentimen suatu kata.

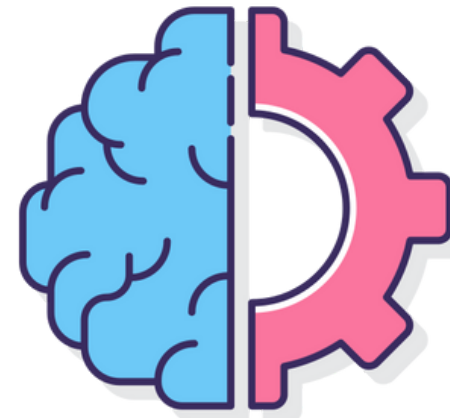
Metode Klasifikasi Menggunakan Machine Learning Model

08.

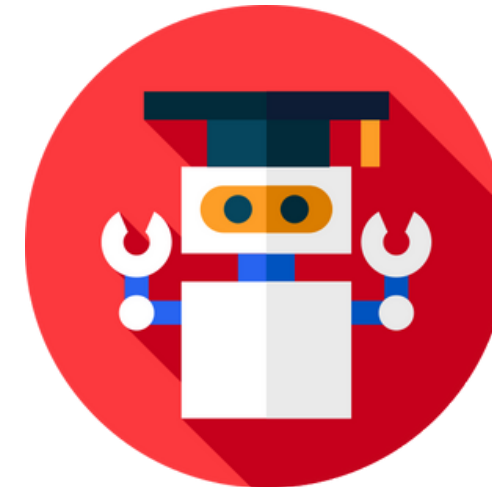
Dataset yang telah diberikan label

Loves the German bakeries in Sydney. Together with my imported honey it feels like home	Positive
@VivaLaLauren Mine is broken too! I miss my sidekick	Negative
Finished fixing my twitter...I had to unfollow and follow everyone again	Negative
@DinahLady I too, liked the movie! I want to buy the DVD when it comes out	Positive
@frugaldougal So sad to hear about @OscarTheCat	Negative
@Mofette brilliant! May the fourth be with you #starwarsday #starwars	Positive
Good morning thespians a bright and sunny day in UK, Spring at last	Positive
@DowneyisDOWNEY Me neither! My laptop's new, has dvd burning/ripping software but I just can't copy the files somehow!	Negative

Dataset ditraining
pada model machine learning



Model machine learning
telah ditraining dan siap digunakan



Hasil Prediksi Sentimen

Prediction
"Positive"

Test Dataset
"Great Service and Affordable
Price"

09.

Kita akan menyelami salah satu metode leksikal

**VADER
(Valence Aware Dictionary
and sEntiment Reasoner)**



[HTTPS://WWW.GREENSCENE.CO.ID/2018/11/27/STAR-WARS-UNGKAP-KEKUATAN-TERBESAR-DARTH-VADER/](https://www.greenscene.co.id/2018/11/27/STAR-WARS-UNGKAP-KEKUATAN-TERBESAR-DARTH-VADER/)

VADER

10.

- Merupakan singkatan dari (**Valence Aware Dictionary and sEntiment Reasoner**)
- **VADER** adalah model yang digunakan untuk menganalisis sentimen yang secara spesifik disesuaikan dengan sentimen yang diungkapkan di **media sosial**
- VADER diciptakan oleh seorang peneliti yang bernama C.J. Hutto
- VADER menggunakan pendekatan **human centric** yang mengombinasikan analisis kualitatif dan validasi yang empiris menggunakan logika manusia dalam penilaian setiap kata dan memanfaatkan konsep **wisdom of crowd**
- **Wisdom of Crowd** bersandar pada konsep bahwa pengetahuan yang kolektif akan lebih bisa dipercaya dibandingkan dengan pengetahuan individual. Perumusan kolektif tersebut tidak dilakukan secara terorganisir melainkan terotomatisasi dan tiap entitasnya saling independen



11.

Contoh Implementasi VADER

- VADER menghasilkan 4 parameter , **compound**, **negatif**, **netral**, dan **positif**.
- Namun hanya **compound** yang biasa digunakan untuk analisis sentimen oleh kebanyakan peneliti

```
# Inisialisasi Library VADER
nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer
sid = SentimentIntensityAnalyzer()

# Menghitung Skor Sentimen suatu Teks
contoh = "This was a great movie"
sid.polarity_scores(contoh)
```

```
{'compound': 0.6249, 'neg': 0.0, 'neu': 0.423, 'pos': 0.577}
```

12.

RP

Mari Kita Membahas Aspek yang Lebih Teknikal

Bagaimana VADER Bisa Menentukan Nilai Valensi Suatu Kata?

13.

- Penilaian dilakukan oleh 10 orang yang independen. 10 orang tersebut melakukan penilaian terhadap 9000 kata yang disediakan.
- Setiap kata dinilai dengan skala dari -4 yang berarti sangat negatif, 4 yang berarti sangat positif, dan 0 yang berarti netral.
- Hasil penilaian setiap kata yang telah dilakukan oleh 10 orang tersebut disimpan pada suatu list.
- Lalu dihitung rata-rata 10 nilai tersebut. Hasil rata-rata tersebut merupakan nilai valensi

Good = [2, 1, 1, 3, 2, 4, 2, 2, 1, 1]

$$\text{Nilai Valensi} = \frac{2 + 1 + 1 + 3 + 2 + 4 + 2 + 2 + 1 + 1}{10} = 1.9$$



Lalu Untuk Menentukan Sentimen 1 Kalimat

14.

- Skor sentimen suatu kalimat dapat ditentukan dengan menyumlahkan seluruh nilai valensi tiap kata pada kalimat tersebut
- Namun hasil skor sentimen tiap kalimat memiliki rentang -1 hingga 1, bukan -4 hingga 4
- Hal ini terjadi karena nilai sentimen tersebut dinormalisasi
- Rumus normalisasi yang digunakan oleh C.J. Hutto :

$$\frac{x}{\sqrt{x^2 + \alpha}}$$

x: Hasil penjumlahan nilai sentimen suatu kalimat

alpha: Parameter normalisasi, diset dengan nilai 15



15.

RP

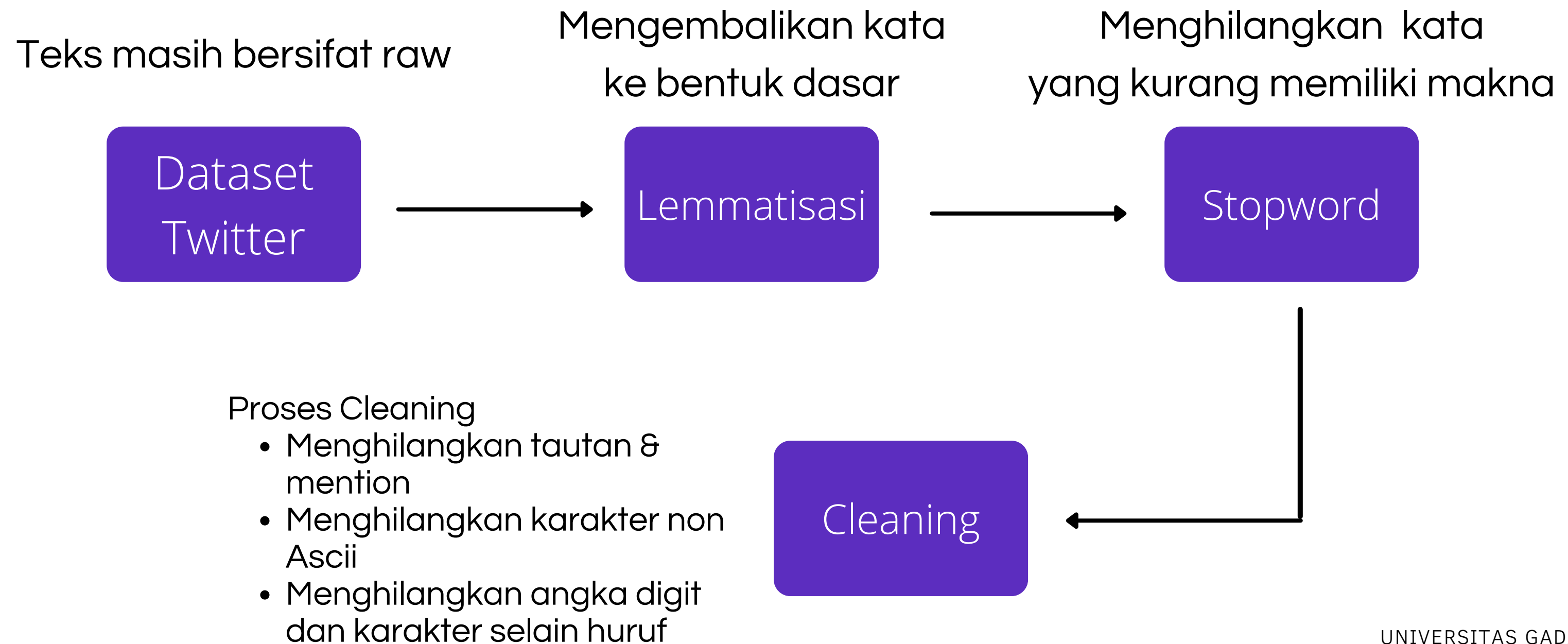
Implementasi VADER pada Dataset Twitter

Overview Dataset

- Menggunakan dataset Twitter yang didapatkan dari Kaggle.
- Dataset terdiri dari kolom **Tweet** dan **Category**.
- **Category** merupakan sentimen yang berisi nilai -1 yang berarti negatif, 0 yang berarti netral, dan 1 yang berarti positif
- Dataset terdiri dari 1630 baris

	tweet	category
0	Wow. Yall needa step it up @Apple RT @heynyla:...	-1
1	What Happened To Apple Inc? http://t.co/FJEX...	0
2	Thank u @apple I can now compile all of the pi...	1
3	The oddly uplifting story of the Apple co-foun...	0
4	@apple can i exchange my iphone for a differen...	0

Pre-Processing Dataset



Dataset Setelah Pre-Processing

	tweet	tweet_lemma	tweet_without_stopwords	tweet_clean
0	when modi promised "minimum government maximum...	when modi promise "minimum government maximum ...	modi promise "minimum government maximum gover...	modi promise minimum government maximum govern...
1	talk all the nonsense and continue all the dra...	talk all the nonsense and continue all the dra...	talk nonsense continue drama vote modi	talk nonsense continue drama vote modi
2	what did just say vote for modi welcome bjp t...	what do just say vote for modi welcome bjp tel...	say vote modi welcome bjp tell rahul main camp...	say vote modi welcome bjp tell rahul main camp...
3	asking his supporters prefix chowkidar their n...	ask his supporters prefix chowkidar their name...	ask supporters prefix chowkidar name modi grea...	ask supporters prefix chowkidar name modi grea...
4	answer who among these the most powerful world...	answer who among these the most powerful world...	answer among powerful world leader today trump...	answer among powerful world leader today trump...

Keterangan kolom:

- tweet: Teks dataset yang masih raw
- tweet_lemma: Teks dari kolom tweet yang telah dilematisasi
- tweet_without_stopwords: Teks yang telah dihilangkan kata-kata yang kurang bermakna
- tweet_clean: Teks yang telah selesai di pre-processing

Implementasi VADER pada Dataset

19.

```
def func_intensity_analyser(review, threshold=0.05, verbose=False):
    sid=SentimentIntensityAnalyzer()
    scores=sid.polarity_scores(review)

    agg_score=scores['compound']
    final_sentiment='POSITIVE' if agg_score>threshold else "NEUTRAL" if agg_score==threshold else
    'NEGATIVE'
    if verbose:
        positive = round(scores['pos'],2)*100
        final= round(agg_score,2)*100
        negative=round(scores['neg'],2)*100
        neutral =round(scores['neu'],2)*100

        print(f'POSITIVE : {positive}%')
        print(f'NEGATIVE : {negative}%')
        print(f'NEUTRAL : {neutral}%')
        print(f'FINAL : {final}%')
        print(f'SENTIMENT : {final_sentiment}')

    return final_sentiment
```

- Tweet yang telah dilakukan Pre-Processing selanjutnya ditentukan sentimennya
- Sentimen ditentukan berdasarkan nilai Compound suatu kalimat

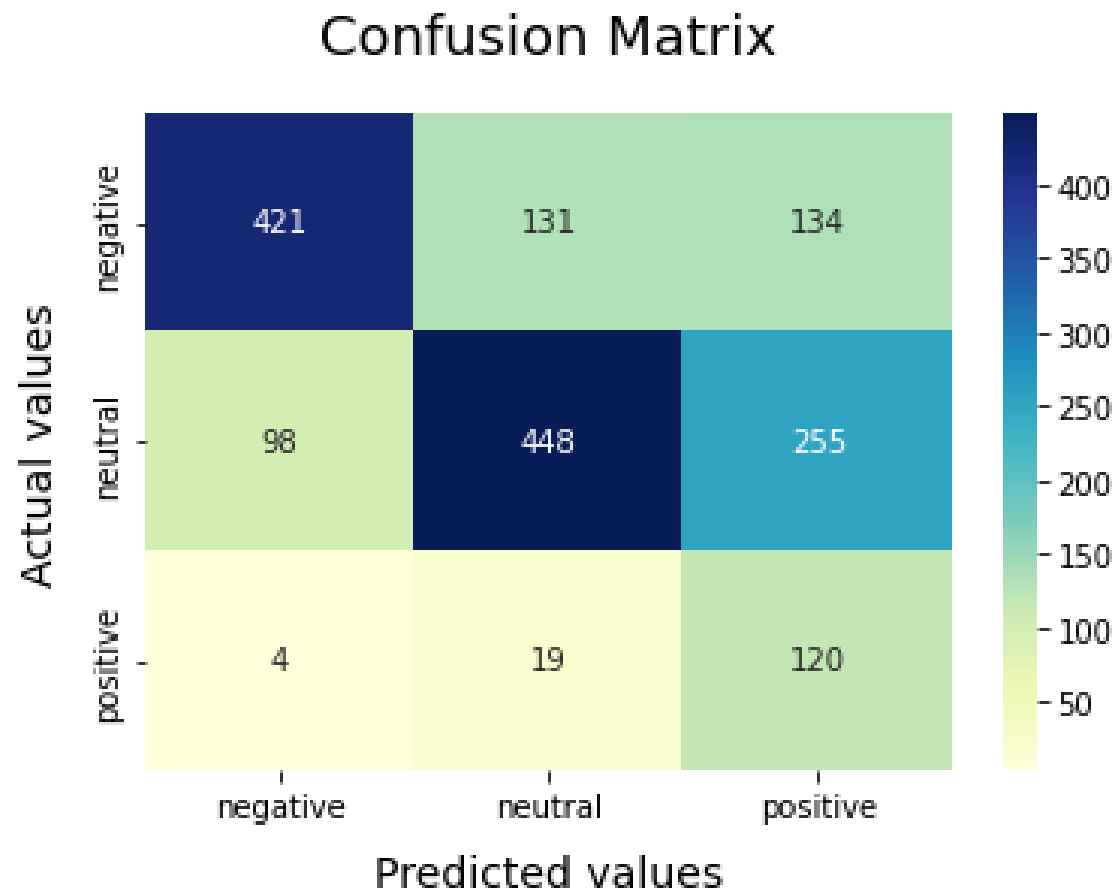
Sentimen Positif: Compound > 0.05

Sentimen Netral: Compound = 0.05

Sentimen Negatif: Compound < 0.05

Performa Model

20.



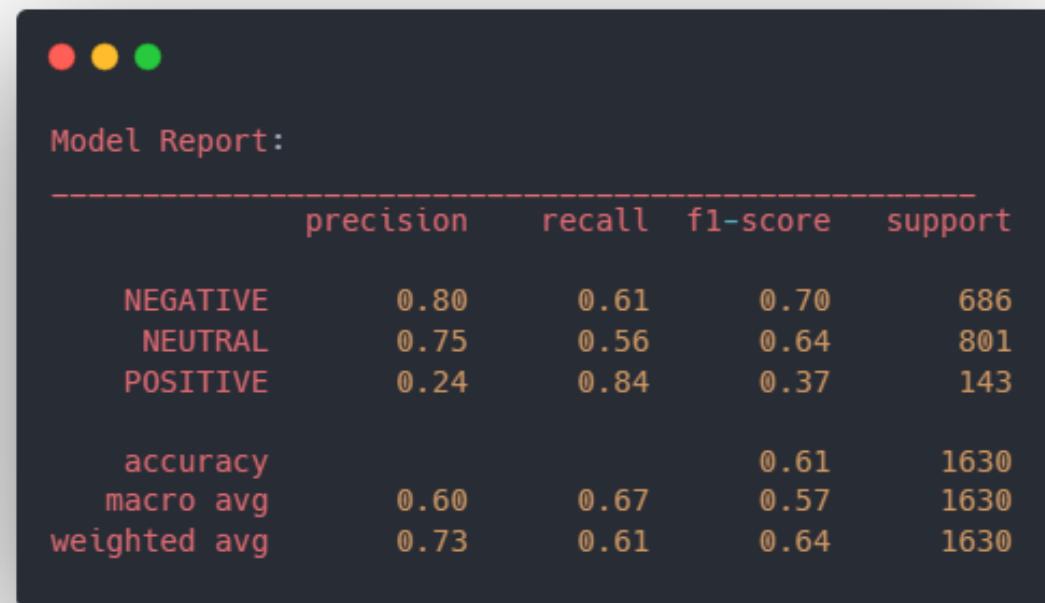
```
Model Report:
```

	precision	recall	f1-score	support
NEGATIVE	0.80	0.61	0.70	686
NEUTRAL	0.75	0.56	0.64	801
POSITIVE	0.24	0.84	0.37	143
accuracy			0.61	1630
macro avg	0.60	0.67	0.57	1630
weighted avg	0.73	0.61	0.64	1630

Performa model menunjukkan hasil akurasi sebesar 0.61

Perbandingan dengan Metode TextBlob

21.

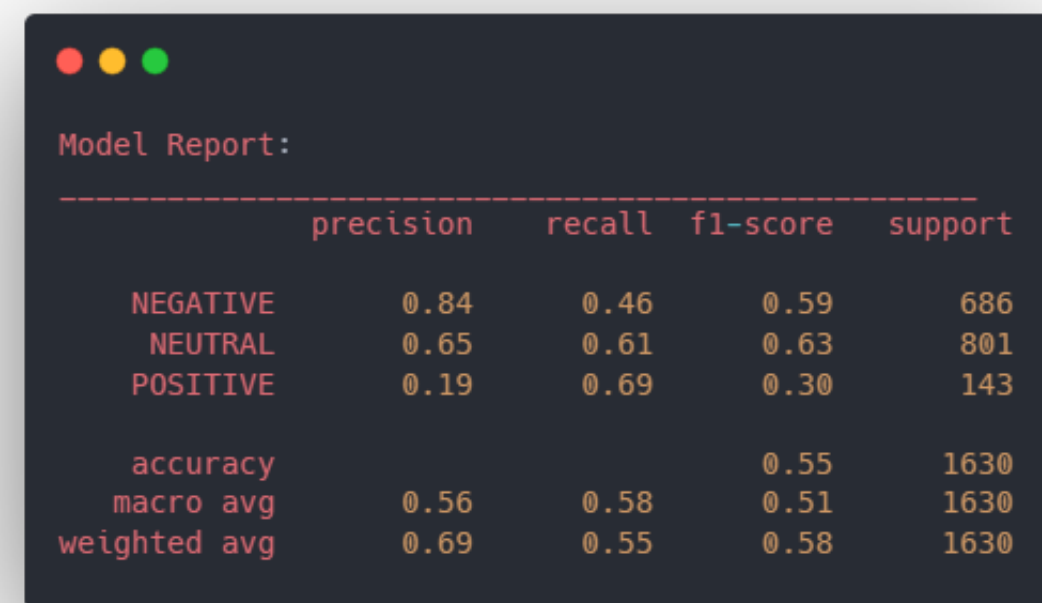


```
Model Report:
-----
              precision    recall  f1-score   support

   NEGATIVE      0.80      0.61      0.70      686
    NEUTRAL      0.75      0.56      0.64      801
    POSITIVE      0.24      0.84      0.37      143

 accuracy      0.61
macro avg      0.60      0.67      0.57      1630
weighted avg   0.73      0.61      0.64      1630
```

Akurasi VADER: 0.61



```
Model Report:
-----
              precision    recall  f1-score   support

   NEGATIVE      0.84      0.46      0.59      686
    NEUTRAL      0.65      0.61      0.63      801
    POSITIVE      0.19      0.69      0.30      143

 accuracy      0.55
macro avg      0.56      0.58      0.51      1630
weighted avg   0.69      0.55      0.58      1630
```

Akurasi TextBlob: 0.55

VADER menunjukkan akurasi yang lebih tinggi dibandingkan TextBlob

22.

RP

Mengapa VADER Memiliki Akurasi yang Lebih Baik?

Kelebihan yang Diberikan VADER

Tanda Baca

- Munculnya tanda baca dapat mengubah sentimen suatu teks. Khususnya tanda seru (!).
- Tanda seru dapat mengubah nilai intensitas sentimen tanpa merubah orientasi semantiknya
- Contoh:
 - "Spiderman No Way Home was great!!!!" memiliki intensitas yang lebih besar dibandingkan "Spiderman No Way Home was great"

Huruf Kapital

- Jika terdapat suatu kata yang seluruh hurufnya kapital di dalam kalimat, sedangkan kata yang lain tidak kapital, nilai intensitas sentimen kalimat tersebut akan meningkat
- Contoh:
 - "Spiderman No Way Home was GREAT!" memiliki intensitas yang lebih besar dibandingkan "Spiderman No Way Home was great!"

Degree Modifiers

- *Degree Modifier* merupakan kata yang meningkatkan derajat suatu kata sifat.
- *Degree Modifier* dapat meningkatkan ataupun menurunkan intensitas sentimen. Tergantung jenis katanya
- Contoh:
 - "This Food is extremely great" memiliki intensitas yang lebih tinggi dibandingkan "This food is great"

Konjungsi "but"

- Kehadiran kata *but* pada suatu kalimat akan memberikan konteks yang kontradiktif antara pernyataan sebelum kata *but* dan yang sesudahnya
- Contoh:
 - "The UI of this app is great, but bug is still occur in this app"
- Kalimat tersebut memiliki sentimen campuran, namun pernyataan setelah kata *but* memberikan intensitas yang lebih dominan

Identifikasi Negasi

- Seringkali suatu kalimat memiliki kata negasi di dalamnya
- Kata negasi tersebut dalam 90% kasus yang C.J. Hutto temukan, mengakibatkan polaritas suatu teks berubah.
- Contoh:
 - "This Game isn't really all that great"

24.

Terima Kasih

Link Kode Program:

<https://colab.research.google.com/drive/1LCt5pgVamtx8338-rVsEFEbrv1sYURif?usp=sharing>

