

Analisis Sentimen menggunakan Naive Bayes (classifier) Untuk Ulasan Mobil

Listrik

Kelompok 8



Rizky Cahya Zuliyanto (3122522018)

Tolak Ivandi Yusditiawan(3122522020)

Yusuf Rahmadhani Asy'Ari (3122522030)

D3 PSDKU Sumenep

PRODI D3 TEKNIK INFORMATIKA

DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER

PENS PSDKU SUMENEP

Proses yang di gunakan untuk Pembangunan model naïve bayes classifier pada tugas ini sebagai berikut :

1. Proses Data Understanding

pemahaman mendalam terhadap data sehingga dapat membuat keputusan yang tepat pada tahap-tahap berikutnya

2. Proses Text Preprocessing

transformasi teks mentah ke dalam format yang lebih bersih dan terstruktur untuk memudahkan analisis lebih lanjut.

3. Proses Visualization Text

Tahap ini untuk menampilkan data teks dalam format visual yang mudah dipahami.

4. Proses Preparation Data

Tahap ini melibatkan pembersihan, transformasi, dan pengorganisasian data mentah menjadi format yang siap untuk analisis lebih lanjut.

5. Proses Modeling

tahap dalam proses analisis data atau pembelajaran mesin di mana model statistik atau algoritma pembelajaran mesin dibangun dan disesuaikan untuk membuat prediksi atau memberikan wawasan berdasarkan data yang telah dipersiapkan.

6. Proses Testing

Tahap ini untuk mengevaluasi kinerja model secara objektif dan memastikan bahwa model dapat menggeneralisasi dengan baik pada data baru.

```
# !pip install Wordcloud
# !pip install Sastrawi

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import re
import nltk
from nltk.corpus import stopwords
from wordcloud import WordCloud, STOPWORDS
from sklearn.feature_extraction.text import CountVectorizer
from nltk.tokenize import RegexpTokenizer

from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report
from sklearn.naive_bayes import MultinomialNB

from wordcloud import WordCloud, STOPWORDS
```

Skrrip Python menggunakan pustaka seperti Wordcloud, Sastrawi, Pandas, NumPy, Matplotlib, Seaborn, NLTK, dan Scikit-Learn. Ini digunakan untuk memproses teks, membuat visualisasi, dan mengembangkan model pembelajaran mesin.

```
from google.colab import drive
drive.mount('/content/drive/')

Drive already mounted at /content/drive/; to attempt to forcibly remount, call drive.mount("/content/drive/", force_remount=True).
```

Ini adalah kode untuk menghubungkan Google Colab dengan Google Drive. Ini memungkinkan akses ke file yang ada di Google Drive dari dalam lingkungan Google Colab.

```
data = pd.read_csv('/content/drive/My Drive/kecerdasan buatan/data/mobil_listrik.csv')
data.head()
```

	id_komentar	nama_akun	tanggal	text_cleaning	sentimen
0	UgzblI5eyrly3-gdUUJ4AaABAg	Sqn Ldr	2023-08-06 12:54:49+00:00	saran sih bikin harga ionic sama kayak brio ...	positif
1	UgzEDUIV3OTrV943p8p4AaABAg	lushen ace	2023-08-04 12:16:23+00:00	problem subsidi kualitas diturunin harga dinai...	negatif
2	UgwqJqu6JMF4EH2CsVV4AaABAg	Fatih Al-Ayyubi	2023-08-04 10:17:57+00:00	baik kualitas kembang dulu baik kualitas motor...	positif
3	UgyYicCMR1rKwuOj2Y14AaABAg	yp office	2023-08-04 08:29:54+00:00	model jelek kwalitas buruk harga mahal croot	negatif
4	UgxKAcLuAwZOQK6es-x4AaABAg	Lembur Kuring	2023-08-04 07:55:37+00:00	syarat ngaco woy anak muda blom punya ruma...	negatif

Next steps: [View recommended plots](#)

Kode ini mencoba untuk membaca file CSV yang terletak di Google Drive dan kemudian menampilkan lima baris pertama dari dataframe yang dibaca. Atribut yang digunakan adalah `text_cleaning` dan `sentimen`.

PROSES DATA UNDERSTANDING

```
data.dtypes
```

id_komentar	object
nama_akun	object
tanggal	object
text_cleaning	object
sentimen	object
dtype:	object

Perintah untuk melihat tipe data dari setiap kolom. Semua tipe data yang ditampilkan yaitu bertipe object

```
data.shape
```

(1517, 5)

Perintah Ini akan memberikan informasi tentang jumlah baris dan kolom dalam dataframe.

```
data = data.dropna()
data.isnull().sum()
```

id_komentar	0
nama_akun	0
tanggal	0
text_cleaning	0
sentimen	0
dtype:	int64

Perintah ini untuk menghapus baris yang mengandung nilai-nilai yang hilang (NaN) dari dataframe dan kemudian memeriksa apakah masih ada nilai-nilai yang hilang setelah penghapusan.

```
[9] data.duplicated().sum()
```

0

Perintah ini memeriksa apakah ada data yang duplikat atau sama dengan data yang lain.

```
data.describe()
```

	id_komentar	nama_akun	tanggal	text_cleaning	sentimen
count	1514	1514	1514	1514	1514
unique	1514	1513	1511	1513	3
top	UgzblI5eyrly3-gdUUJ4AaABAg	Albert	2023-06-06 14:02:09+00:00	mahal	negatif
freq	1	2	2	2	868

Perintah digunakan untuk memberikan ringkasan statistik dari dataframe, termasuk ringkasan statistik untuk kolom-kolom numerik seperti mean, median, minimum, maksimum, dan quartile.

PROSES TEXT PREPROCESSING

```
data.head()
```

	id_komentar	nama_akun	tanggal	text_cleaning	sentimen
0	Ugzblf5eyrly3-gdUUJ4AaABAg	Sqn Ldr	2023-08-06 12:54:49+00:00	saran sih bikin harga ionic sama kayak brio ...	positif
1	UgzEDUIV3OTrV943p8p4AaABAg	lushen ace	2023-08-04 12:16:23+00:00	problem subsidi kualitas diturunin harga dinai...	negatif
2	UgwwJqu6JMF4EH2CsVV4AaABAg	Fatih Al-Ayyubi	2023-08-04 10:17:57+00:00	baik kualitas kembang dulu baik kualitas motor...	positif
3	UgyYicCMR1rKwuOj2Y14AaABAg	yp office	2023-08-04 08:29:54+00:00	model jelek kwalitas buruk harga mahal croot	negatif
4	UgxKAcLuAwZQOK6es-x4AaABAg	Lembur Kuring	2023-08-04 07:55:37+00:00	syarat ngaco woy anak muda blom punya ruma...	negatif

Perintah untuk memanggil 5 data teratas

```
data = data.drop(columns='id_komentar')
data.head()
```

	nama_akun	tanggal	text_cleaning	sentimen
0	Sqn Ldr	2023-08-06 12:54:49+00:00	saran sih bikin harga ionic sama kayak brio ...	positif
1	lushen ace	2023-08-04 12:16:23+00:00	problem subsidi kualitas diturunin harga dinai...	negatif
2	Fatih Al-Ayyubi	2023-08-04 10:17:57+00:00	baik kualitas kembang dulu baik kualitas motor...	positif
3	yp office	2023-08-04 08:29:54+00:00	model jelek kwalitas buruk harga mahal croot	negatif
4	Lembur Kuring	2023-08-04 07:55:37+00:00	syarat ngaco woy anak muda blom punya ruma...	negatif

Perintah untuk menghapus colum id_komentar pada data, dikarena id_komentar tidak di butuh kan pada proses ini.

```
# mengubah teks menjadi lower
data['text_cleaning'] = data['text_cleaning'].str.lower()
data.head()
```

	nama_akun	tanggal	text_cleaning	sentimen
0	Sqn Ldr	2023-08-06 12:54:49+00:00	saran sih bikin harga ionic sama kayak brio ...	positif
1	lushen ace	2023-08-04 12:16:23+00:00	problem subsidi kualitas diturunin harga dinai...	negatif
2	Fatih Al-Ayyubi	2023-08-04 10:17:57+00:00	baik kualitas kembang dulu baik kualitas motor...	positif
3	yp office	2023-08-04 08:29:54+00:00	model jelek kwalitas buruk harga mahal croot	negatif
4	Lembur Kuring	2023-08-04 07:55:37+00:00	syarat ngaco woy anak muda blom punya ruma...	negatif

Perintah untuk mengubah text menjadi lower pada data di colum text_cleaning.

```
# Normalisasi
norm = {" dgn " : " dengan ", " gue " : " saya ", " ngaco " : " sembarangan ", " kwalitas " : " kualitas ", ' blom ' : ' belum ', ' aing ' : ' aku ', ' tehn
def normalisasi(str_text):
    for i in norm:
        str_text = str_text.replace(i, norm[i])
    return str_text
data['text_cleaning'] = data['text_cleaning'].apply(lambda x: normalisasi(x))
```

Perintah ini untuk mengubah huruf atau kata yang tidak terdeteksi pada kamus KBBI menjadi kata yang mudah dipahami atau terdeteksi bertujuan untuk meningkatkan performa model yang di buat.

```
# Stopword
import Sastrawi
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory, StopWordRemover, ArrayDictionary

more_stop_word = []

stop_words = StopWordRemoverFactory().get_stop_words()
new_array = ArrayDictionary(stop_words)
stop_words_remover_new = StopWordRemover(new_array)

def stopword(str_text):
    str_text = stop_words_remover_new.remove(str_text)
    return str_text

data['text_cleaning'] = data['text_cleaning'].apply(lambda x: stopword(x))
data.head()
```

	nama_akun	tanggal	text_cleaning	sentimen
0	Sqn Ldr	2023-08-06 12:54:49+00:00	saran sih bikin harga ionic sama kayak brio ...	positif
1	lushen ace	2023-08-04 12:16:23+00:00	masalah subsidi kualitas diturunin harga dina...	negatif
2	Fatih Al-Ayyubi	2023-08-04 10:17:57+00:00	baik kualitas kembang dulu baik kualitas motor...	positif
3	yp office	2023-08-04 08:29:54+00:00	model jelek kualitas buruk harga mahal croot	negatif

Perintah untuk menghilangkan kata yang tidak terlalu penting dalam proses pembangunan sentimen analisis, maka dari itu kata-kata tersebut harus di hilangkan.

```
# Tokenisasi
tokenized = data['text_cleaning'].apply(lambda x:x.split())
tokenized
```

```
0      [saran, sih, bikin, harga, ionic, sama, kayak,...
1      [masalah, subsidi, kualitas, diturunin, harga,...
2      [baik, kualitas, kembang, dulu, baik, kualitas...
3      [model, jelek, kualitas, buruk, harga, mahal, ...
4      [syarat, sembarangan, woy, anak, muda, punya, ...
      ...
1512   [apa, kabar, padahal, negri, luar, biasa, negara]
1513   [antar, anak, sekolah, antar, bantu, pasar, kl...
1514   [esemka, bangga, solo]
1515   [cerdas, orang, dan, pasar, jalan, x, da, gima...
1516   [niat, beli, ev, murah, malah, ikut, dinaikin,...
Name: text_cleaning, Length: 1514, dtype: object
```

Perintah untuk memisah semua kata pada setiap kalimat yang berada di colum text_cleaning

```
# Stemming
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

def stemming(text_cleaning):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    do = []
    for w in text_cleaning:
        dt = stemmer.stem(w)
        do.append(dt)
    d_clean = []
    d_clean = " ".join(do)
    print(d_clean)
    return d_clean

tokenized = tokenized.apply(stemming)

tokenized.to_csv('/content/drive/MyDrive/kecerdasan buatan/data/mobil_listrik_bersih1.csv', index=False)
data_clean = pd.read_csv('/content/drive/MyDrive/kecerdasan buatan/data/mobil_listrik_bersih1.csv', encoding='latin1')
```

Perintah berguna untuk mengubah beberapa kata-kata menjadi kata dasar. Dan menyimpan data yang di ubah ke dalam drive yang di tentukan

...

>

data ▾

✓

≡

⌵

i

Jenis ▾

Orang ▾

Dimodifikasi ▾



🔊

Pembaruan pada pintasan keyboard Pada Kamis, 01 Agustus 2024, pintasan keyboard Drive akan diperbarui untuk memberi Anda navigasi huruf pertama Pelajari lebih lanjut

✕


Lihat pintasan

Perbarui sekarang


Nama	↓	Terakhi... ▾	⋮
	mobil_listrik.csv	4 Jun 2024	⋮
	mobil_listrik_bersih1.csv	23.49	⋮


Data setelah berhasil di simpan di drive

```
[25] data_clean.head()
```



text_cleaning



0	saran sih bikin harga ionic sama kayak brio in...	
1	masalah subsidi kualitas diturunin harga dinai...	
2	baik kualitas kembang dulu baik kualitas motor...	
3	model jelek kualitas buruk harga mahal croot	
4	syarat sembarang woy anak muda punya rumah jad...	

Perintah untuk menampilkan 5 data di awal.

```
at1 = pd.read_csv('/content/drive/MyDrive/kecerdasan buatan/data/mobil_listrik_bersih1.csv')
at2 = pd.read_csv('/content/drive/MyDrive/kecerdasan buatan/data/mobil_listrik.csv')
att2 = at2['sentimen']

data_clean = pd.concat([at1, att2], axis=1)
data_clean
```

	text_cleaning	sentimen
0	saran sih bikin harga ionic sama kayak brio in...	positif
1	masalah subsidi kualitas diturunin harga dinai...	negatif
2	baik kualitas kembang dulu baik kualitas motor...	positif
3	model jelek kualitas buruk harga mahal croot	negatif
4	syarat sembarang woy anak muda punya rumah jad...	negatif

Perintah menggabungkan dua data agar bisa menampilkan text_cleaning dan sentimen.

```
[28] # Menghapus baris dengan nilai teks yang hilang (np.nan)
data_clean = data_clean.dropna()

data_clean = data_clean[data_clean['sentimen'] != 'netral']
data_clean
```

	text_cleaning	sentimen
0	saran sih bikin harga ionic sama kayak brio in...	positif
1	masalah subsidi kualitas diturunin harga dinai...	negatif
2	baik kualitas kembang dulu baik kualitas motor...	positif
3	model jelek kualitas buruk harga mahal croot	negatif
4	syarat sembarang woy anak muda punya rumah jad...	negatif
...
1509	apa kabar padahal negri luar biasa negara	negatif
1510	antar anak sekolah antar bantu pasar klw jauh ...	positif
1511	esemka bangga solo	positif

Perintah ini menghapus data yang memiliki sentimen netral dikarena model ini hanya menggunakan sentimen positif dan negatif

```
data_clean = data_clean.replace({'positif':1, 'negatif':0})
data_clean.head()
```

	text_cleaning	sentimen
0	saran sih bikin harga ionic sama kayak brio in...	1
1	masalah subsidi kualitas diturunin harga dinai...	0
2	baik kualitas kembang dulu baik kualitas motor...	1
3	model jelek kualitas buruk harga mahal croot	0
4	syarat sembarang woy anak muda punya rumah jad...	0

Dikarenakan project menggunakan metode Naive Bayes, maka sentimen positif dan negatif harus di ubah menjadi nilai atau angka. Perintah di atas berguna mengubah sentimen positif dan negatif menjadi nilai 1 dan 0.

PROSES VISUALISASI KATA

```
data_negatif = data_clean[data_clean['sentimen'] == 0]
data_positif = data_clean[data_clean['sentimen'] == 1]
```

Perintah untuk menyaring DataFrame berdasarkan nilai-nilai 'sentimen' yang mewakili sentimen negatif dan positif.

```
all_text_s0 = ' '.join(word for word in data_negatif["text_cleaning"])
wordcloud = WordCloud(colormap='Reds', width=1000, height=1000, mode='RGBA', background_color='white').generate(all_text_s0)
plt.figure(figsize=(9, 6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title("Visualisasi Kata Negatif")
plt.margins(x=0, y=0)
plt.show()
```



Perintah untuk memvisualisasi kata negative pada data yang di jalan kan.

```
all_word_text_cell['.join(word for word in data_positif["text_cleaning"])']
wordcloud = WordCloud(colormap='Blues', width=1000, height=1000, mode="RGBA", background_color='white').generate(all_text_s1)
plt.figure(figsize=(9, 6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.title("Visualisasi Kata Positif")
plt.margins(x=0, y=0)
plt.show()
```

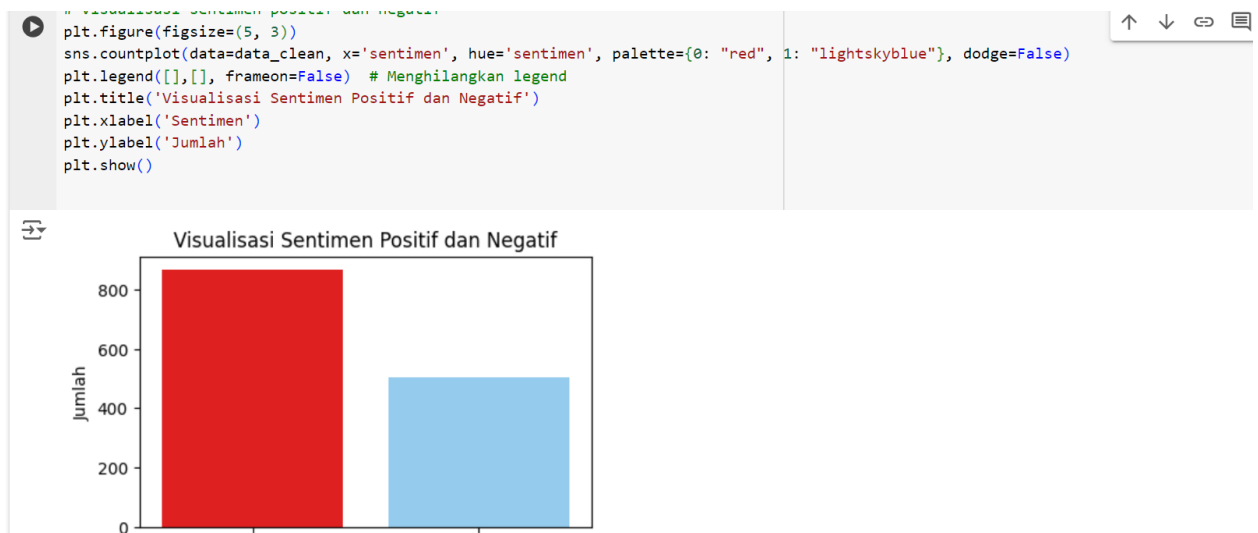


Perintah untuk memvisualisasi kata positif pada data yang di jalankan.

```
data_clean['sentimen'].value_counts()
```

```
sentimen
0      868
1      503
Name: count, dtype: int64
```

perintah untuk mengetahui jumlah sentiment positif dan negative.



Perintah ini berguna untuk menentukan diagram sentimen positif dan negative.

PROSES DATA PREPARATION DATA

```
# Memisahkan data menjadi data latih dan data uji
X = data_clean['text_cleaning']
y = data_clean['sentimen']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.15, random_state=42)
```

Perintah ini berguna untuk memisahkan data latih dan data uji.

```
# Menggunakan CountVectorizer untuk mengubah teks menjadi fitur numerik
vectorizer = CountVectorizer()
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)
```

Perintah `countvectorizer` berguna untuk mengubah teks menjadi fitur menarik.

```
[55] # Melakukan oversampling menggunakan SMOTE pada data latih
smote = SMOTE(random_state=42)
X_train_resampled, y_train_resampled = smote.fit_resample(X_train_vec, y_train)
```

Perintah untuk melakukan oversampling menggunakan SMOTE pada data latih.



Contoh data hasil resampling.

PROSES MODELING

```

[57] naive_bayes = MultinomialNB()
naive_bayes.fit(X_train_resampled, y_train_resampled)

```

▼ MultinomialNB
MultinomialNB()

Perintah untuk inialisasi model naïve bayes classifier.

```

y_pred = naive_bayes.predict(X_test_vec)

```

Perintah membuat variable y dan memasukan variabel naïve bayes untuk memprediksi x.

```

# Evaluasi Model
accuracy = accuracy_score(y_test, y_pred)
classification_rep = classification_report(y_test, y_pred, target_names=['negatif', 'positif'])

print("Akurasi Model Naive Bayes : ", accuracy)
print("\nLaporan Klasifikasi :\n", classification_rep)

```

Akurasi Model Naive Bayes : 0.7766990291262136

Laporan Klasifikasi :

	precision	recall	f1-score	support
negatif	0.80	0.89	0.84	137
positif	0.72	0.55	0.62	69
accuracy			0.78	206
macro avg	0.76	0.72	0.73	206
weighted avg	0.77	0.78	0.77	206

Perintah untuk mengetahui Gambaran evaluasi dari model yang di buat atau di bangun.

TESTING

Melakukan testing terhadap model yang telah di buat tadi:

```
new_text = input("\nMasukkan teks baru: ")
new_text_vec = vectorizer.transform([new_text])
predicted_sentimen = naive_bayes.predict(new_text_vec)

if predicted_sentimen[0] == 1:
    sentiment_label = "positif"
elif predicted_sentimen[0] == 0:
    sentiment_label = "negatif"

print("Hasil Analisis Sentimen untuk Teks Baru :", sentiment_label)
```

Masukkan teks baru: kualitas bagus, performa mantap
Hasil Analisis Sentimen untuk Teks Baru : positif

disini saat memasukan data kualitas bagus, permorma mantap otomatis akan terdeteksi bahwa itu sentiment positif

```
new_text = input("\nMasukkan teks baru: ")
new_text_vec = vectorizer.transform([new_text])
predicted_sentimen = naive_bayes.predict(new_text_vec)

if predicted_sentimen[0] == 1:
    sentiment_label = "positif"
elif predicted_sentimen[0] == 0:
    sentiment_label = "negatif"

print("Hasil Analisis Sentimen untuk Teks Baru :", sentiment_label)
```

Masukkan teks baru: mobil rusak dan berkarat
Hasil Analisis Sentimen untuk Teks Baru : negatif

Demikian juga sebalik nya jika memasukan teks mobil rusak dan berkarat otomatis model akan mendeteksi bahwa itu adalah sentiment negative.

Demikian proses sentiment analisis menggunakan model naïve bayes terhadap ulasan mobil Listrik di Indonesia.