

NAMA : MOHAMMAD REZA  
NIM : A11.2022.14741  
MATA KULIAH : DATA MINNING

## DRAFT EKSPERIMEN TUGAS AKHIR

- A. **TOPIK ATAU JUDUL:** Analisis Sentimen Komentar Youtube menggunakan Algoritma Klasifikasi Naïve Bayes
- B. **DESKRIPSI SINGKAT ( BUSSINES UNDERSTANDING/ DATA UNDERSTANDING ):**  
Dalam eksperimen ini, saya akan menerapkan metode klasifikasi seperti naïve bayes untuk menganalisis sentimen dari komentar pengguna di platform YouTube. Tujuan dari eksperimen ini adalah untuk mengidentifikasi sentimen dalam komentar, apakah positif, negatif, atau netral.
- C. **TUJUAN:**
- Masalah: Bagaimana cara mengklasifikasikan sentimen dalam komentar YouTube menjadi kategori positif, negatif, atau netral?
  - Tujuan: Menggunakan algoritma klasifikasi naïve bayes untuk mengidentifikasi dan menganalisis sentimen pada komentar YouTube serta mengevaluasi kinerja model dalam mengklasifikasikan sentimen.
- D. **ALUR / TAHAPAN / KERANGKA EKSPERIMEN**
- **BU ( BUSSINES UNDERSTANDING ):** Mengidentifikasi tujuan analisis sentimen untuk memahami pandangan penonton terhadap konten di YouTube.
  - **DU ( DATA UNDERSTANDING ):** Mengumpulkan dataset komentar YouTube. Dataset ini bisa diperoleh melalui scraping komentar dari video tertentu atau menggunakan dataset komentar YouTube yang tersedia di platform seperti Kaggle.
  - **Pre-proccesing Data:** Membersihkan teks komentar dari simbol, emoji, URL, dan karakter khusus, Memberikan StopWords, Stemming.
  - **Modelling:** Menggunakan algoritma klasifikasi seperti **Naïve Bayes**, SVM, atau Logistic Regression untuk menentukan sentimen komentar (positif, negatif, atau netral).
  - **Evaluasi:** Mengukur akurasi model dengan menggunakan metrik seperti confusion matrix.
  - **Implementasi:** Menyajikan hasil dari analisis sentimen dan memberikan kesimpulan mengenai sentimen yang terdeteksi.
- E. **PENJELASAN DATASET**
- **DATASET:** Dataset yang digunakan dalam eksperimen ini terdiri dari kumpulan komentar dari platform YouTube. Data ini dapat diperoleh melalui web scraping, API Youtube atau menggunakan dataset yang sudah tersedia di Kaggle.
  - **ATTRIBUTE UTAMA:**
    - a. Comment: Beirisi komentar oengguna
    - b. Sentiment: Label sentiment ( jika tidak ada label pada dataset maka bisa dilakukan proses labeling terlebih dahulu ).

## Eksperimen:

```
#Libraries
!pip install deep-translator
!pip install tqdm
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import os

# Import functions for data preprocessing & data preparation
from sklearn.preprocessing import LabelEncoder
from sklearn.utils import resample
from sklearn.feature_extraction.text import CountVectorizer
from nltk.sentiment.vader import SentimentIntensityAnalyzer

from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from nltk.stem import PorterStemmer, LancasterStemmer
from nltk.stem.snowball import SnowballStemmer
from nltk.corpus import stopwords
from nltk.corpus import wordnet
import string
from string import punctuation
import nltk
import re
```

```
data = pd.read_csv('input/pemiluyt.csv')
data.columns
data1=data.drop(['Timestamp', 'VideoID', 'Date'],axis=1)
data1
```

	Username	Comment
0	@surveyorindonesia8772	Semoga Amanah.. Aamiin YRA, 🙏
1	@delis1247	KAMI MEMINTA PEMILIHAN ULANG. KAMI MINTA KEA...
2	@lalaaulia2451	AMIN❤️🙏
3	@uyikwer1502	01
4	@ubaidillah1900	Udah langsung aja 01 menang 56%
5	@FirdausiJuliardha	I like Number Two, Number Two...
6	@baiqnovievijaya5705	👍👍👍👍👍👍👍👍
7	@user-dn1vj8hn2j	01 Amin
8	@user-ln3gy3ie6l	Amin 01

```

from deep_translator import GoogleTranslator
from tqdm import tqdm
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer

# Pastikan untuk mendownload lexicon VADER
nltk.download('vader_lexicon')

# Inisialisasi SentimentIntensityAnalyzer
sentiments = SentimentIntensityAnalyzer()

# Terjemahkan teks komentar ke bahasa Inggris dengan progress bar
data['Comment_EN'] = [GoogleTranslator(source='id', target='en').translate(comment) if comment else "" for comment in tqdm(data['comment'], desc="Translating")]

# Isi None atau NaN dalam Comment_EN dengan string kosong untuk menghindari error
data['Comment_EN'] = data['Comment_EN'].fillna("")

# Analisis sentimen pada teks yang sudah diterjemahkan
data['Positive'] = [sentiments.polarity_scores(i)['pos'] if i else 0 for i in data['Comment_EN']]
data['Negative'] = [sentiments.polarity_scores(i)['neg'] if i else 0 for i in data['Comment_EN']]
data['Neutral'] = [sentiments.polarity_scores(i)['neu'] if i else 0 for i in data['Comment_EN']]
data['Compound'] = [sentiments.polarity_scores(i)['compound'] if i else 0 for i in data['Comment_EN']]

# Mengklasifikasi hasil sentimen
score = data['Compound'].values
sentiment = []
for i in score:
    if i >= 0.05:
        sentiment.append('Positive')
    elif i <= -0.05:
        sentiment.append('Negative')
    else:
        sentiment.append('Neutral')
data['Sentiment'] = sentiment

# Lihat hasilnya
data.head()

```

```

[ ] processed_data = {
    'Sentence':data_copy.Comment,
    'Sentiment':data_copy['Sentiment']
}

processed_data = pd.DataFrame(processed_data)
processed_data.head()

```

	Sentence	Sentiment
0	semoga amanah aamiin ya	Positive
1	kami meminta pemilihan ulang kami minta keadilan	Positive
2	amin	Neutral
3	01	Neutral
4	udah langsung aja 01 menang 56	Positive

Langkah berikutnya: [Buat kode dengan processed\\_data](#) [Lihat plot yang direkomendasikan](#)

```

processed_data['Sentiment'].value_counts()

```

Sentiment	count
Neutral	30
Positive	18
Negative	2

dtype: int64