

UJIAN AKHIR SEMESTER (UAS)

MATA KULIAH DATA MINING



DISUSUN OLEH

M NURHASAN MAHMUDI

241012000090

PROGRAM PASCASARJANA

MAGISTER TEKNIK INFORMATIKA UNIVERSITAS PAMULANG

TANGERANG SELATAN

2024

Daftar Isi

JUDUL.....	2
PENDAHULUAN	3
Latar Belakang	3
Tujuan Penelitian	3
METODE PENELITIAN.....	4
Teknik Pengambilan Data.....	4
HASIL ANALISIS SENTIMEN	8
Widget Digunakan Pada Orange Data Mining	8
Proses Analisis	9
Workflow Orange Data Mining	11
KESIMPULAN.....	17

JUDUL

**“ANALISIS SENTIMEN PEMECATAN SHIN TAE-YONG DALAM MENJADI
PELATIH TIMNAS SEPAKBOLA INDONESIA TAHUN 2025”**

PENDAHULUAN

Latar Belakang

Sepak bola merupakan olahraga yang sangat populer di Indonesia, dengan jutaan pendukung yang antusias mengikuti perkembangan Tim Nasional. Penunjukan Shin Tae-yong sebagai pelatih pada tahun 2019 membawa harapan baru bagi peningkatan prestasi tim. Di bawah kepemimpinannya, Indonesia mencapai beberapa pencapaian signifikan, termasuk kemenangan atas Arab Saudi dan lolos ke babak ketiga kualifikasi Piala Dunia 2026.

Namun, pada Januari 2025, PSSI memutuskan untuk mengakhiri kontrak dengan Shin Tae-yong, dengan alasan perlunya kepemimpinan yang lebih efektif dan komunikasi yang lebih baik dalam tim.

Keputusan ini menimbulkan berbagai reaksi dari masyarakat, media, dan pengamat sepak bola, yang mempertanyakan timing dan alasan di balik pemecatan tersebut.

Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap pemecatan Shin Tae-yong sebagai pelatih Tim Nasional Sepak Bola Indonesia pada tahun 2025. Secara khusus, penelitian ini akan:

- Mengidentifikasi pola sentimen (positif, negatif, netral) dalam komentar publik terkait pemecatan tersebut.
- Menganalisis alasan dan argumen yang mendasari sentimen publik.
- Menilai dampak keputusan pemecatan terhadap persepsi publik terhadap PSSI dan Tim Nasional.

METODE PENELITIAN

Penelitian ini menggunakan metode pengumpulan data melalui teknik *web crawling* untuk mengumpulkan komentar dari video pemberitaan yang relevan di platform media sosial dan situs berbagi video. Tahapan metode penelitian meliputi:

- **Pengumpulan Data:** Menggunakan skrip crawling untuk mengidentifikasi dan mengunduh komentar dari video yang membahas pemecatan Shin Tae-yong. Sumber data mencakup *platform* seperti *YouTube*, di mana diskusi publik berlangsung secara aktif.
- **Pra-pemrosesan Data:** Membersihkan data komentar dari elemen non-teks, seperti emoji, tanda baca berlebih, dan informasi meta yang tidak relevan. Selain itu, dilakukan normalisasi teks untuk memastikan konsistensi dalam analisis.
- **Analisis Sentimen:** Menggunakan algoritma *Natural Language Processing* (NLP) untuk mengklasifikasikan komentar ke dalam kategori sentimen positif, negatif, atau netral. Model analisis sentimen dilatih khusus untuk konteks bahasa Indonesia guna meningkatkan akurasi.
- **Analisis Tematik:** Mengidentifikasi tema dan topik utama yang muncul dalam komentar untuk memahami alasan di balik sentimen yang diekspresikan oleh publik. Analisis ini dilakukan dengan pendekatan kualitatif, termasuk pengkodean manual dan penggunaan perangkat lunak *Orange Data Mining*.

Teknik Pengambilan Data

Pengambilan data yang akan di proses pada analisis ini berasal dari komentar dari video youtube dengan judul yang relevan dengan judul pada Analisa ini. Dengan menggunakan *script Python* dan Menggunakan fasilitas dari API Youtube Data API V3 pada Google Cloud Console. Dengan script yang digunakan seperti dibawah ini :

```
pip install google-api-python-client

import googleapiclient.discovery
import csv
import os

# Konfigurasi API
API_KEY = "AIzaSyAmJ9bUmdkQdBV7L3wjg2VuWaSwDFN5mIc"
YOUTUBE_API_SERVICE_NAME = "youtube"
YOUTUBE_API_VERSION = "v3"

# Variabel yang dapat diubah
```

```

MAX_VIDEOS = 15 # Jumlah maksimum video yang diambil
MAX_COMMENTS = 100 # Jumlah maksimum komentar per video

def search_videos(query, max_results=MAX_VIDEOS, published_after="2025-01-05T00:00:00Z", region_code="ID"):
    youtube = googleapiclient.discovery.build(
        YOUTUBE_API_SERVICE_NAME, YOUTUBE_API_VERSION,
        developerKey=API_KEY
    )

    search_response = youtube.search().list(
        q=query,
        part="id,snippet",
        maxResults=max_results,
        type="video",
        publishedAfter=published_after,
        regionCode=region_code
    ).execute()

    videos = []
    for item in search_response.get("items", []):
        video_id = item["id"]["videoId"]
        title = item["snippet"]["title"]
        published_at = item["snippet"]["publishedAt"]
        videos.append({"video_id": video_id, "title": title,
            "published_at": published_at})

    return videos

def get_video_comments(video_id, max_comments=MAX_COMMENTS):
    youtube = googleapiclient.discovery.build(
        YOUTUBE_API_SERVICE_NAME, YOUTUBE_API_VERSION,
        developerKey=API_KEY
    )

    comments = []
    next_page_token = None

    try:
        while len(comments) < max_comments:
            comment_response = youtube.commentThreads().list(
                part="snippet",
                videoId=video_id,
                maxResults=min(100, max_comments - len(comments)),
                pageToken=next_page_token
            ).execute()

            for item in comment_response.get("items", []):
                comment =
item["snippet"]["topLevelComment"]["snippet"]["textDisplay"]
                username =
item["snippet"]["topLevelComment"]["snippet"]["authorDisplayName"]
                comments.append({"comment": comment, "username":
username})

            next_page_token = comment_response.get("nextPageToken")
            if not next_page_token:
                break

```

```

except googleapiclient.errors.HttpError as e:
    if e.resp.status == 403:
        print(f"Komentar dinonaktifkan untuk video ID {video_id}.")
    else:
        print(f"Terjadi kesalahan saat mengambil komentar untuk video ID {video_id}: {e}")

    return comments

if __name__ == "__main__":
    # Keyword pencarian
    query = "Copot Shin Tae-Yong, Copot STY"

    # Cari video
    videos = search_videos(query)

    print("Ditemukan video:")
    for idx, video in enumerate(videos):
        print(f"{idx + 1}. {video['title']} (ID: {video['video_id']})")

    total_comments = 0

    # Simpan komentar ke file CSV
    csv_file_path = "youtube_comments.csv"
    with open(csv_file_path, "w", encoding="utf-8", newline="") as csvfile:
        csvwriter = csv.writer(csvfile)
        # Tulis header CSV
        csvwriter.writerow(["Video Title", "Video ID", "Video Link", "Published Date", "Username", "Comment"])

        # Ambil komentar dari setiap video
        for video in videos:
            print(f"\nMengambil komentar untuk video: {video['title']}")
            comments = get_video_comments(video["video_id"])
            total_comments += len(comments)

            # Tulis komentar ke CSV
            for comment_data in comments:
                csvwriter.writerow([
                    video["title"],
                    video["video_id"],
                    f"https://www.youtube.com/watch?v={video['video_id']}",
                    video["published_at"],
                    comment_data["username"],
                    comment_data["comment"]
                ])

    print(f"\nSemua komentar disimpan ke {csv_file_path}")
    print(f"Total komentar yang diambil: {total_comments}")

    # Auto-download file CSV
    if os.path.exists(csv_file_path):
        print(f"File {csv_file_path} siap untuk diunduh.")

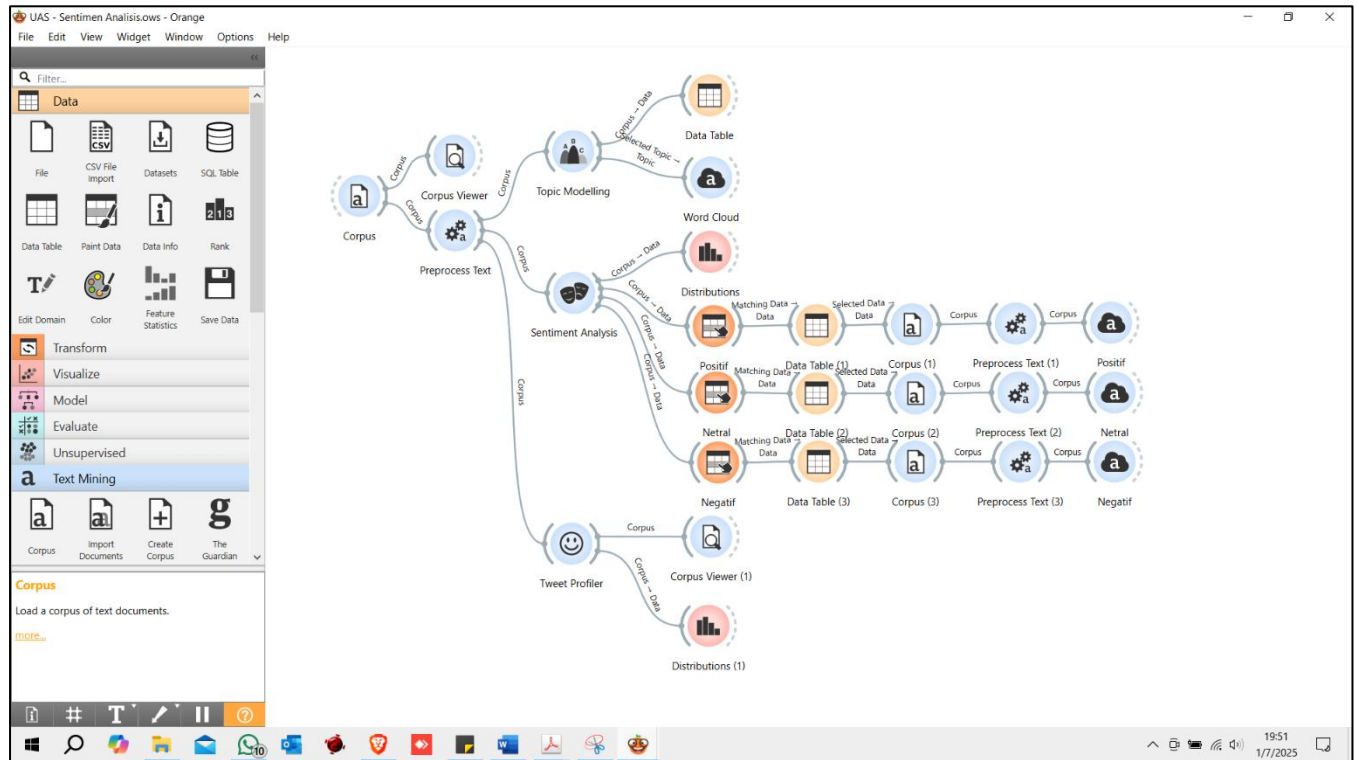
```

Dari *script* tersebut bertujuan untuk mendapatkan 100 komentar dari video yang judulnya relevan dengan *keyword* “Copot Shin Tae-Yong” & “Copot STY”. Dan menghasilkan 1266 data komentar dari 15 Video sebagai Berikut :

	A	B	C	D	E	F	G	H
1	Video Title	Video ID	Video Link	Published Date	Username	Comment		
1260	ðŸŽŸ" LIVE STREAMING	AOWSAf8Tx18	https://www.you	2025-01-06T05:39:21Z	@IrwandiCaniago-h1b	Kita harus yakin PSSI mengam		
1261	ðŸŽŸ" LIVE STREAMING	AOWSAf8Tx18	https://www.you	2025-01-06T05:39:21Z	@udothudayansah6520	Ini inti permasalahannya sbny		
1262	ðŸŽŸ" LIVE STREAMING	AOWSAf8Tx18	https://www.you	2025-01-06T05:39:21Z	@TEKUKUR.KALIMANTAN	Seperti nya Indonesia meman		
1263	ðŸŽŸ" LIVE STREAMING	AOWSAf8Tx18	https://www.you	2025-01-06T05:39:21Z	@Koromogu	Semoga malaysia mengambil		
1264	ðŸŽŸ" LIVE STREAMING	AOWSAf8Tx18	https://www.you	2025-01-06T05:39:21Z	@ischodzakydzaky3106	Jngn pernah pakai pelatih Ind		
1265	ðŸŽŸ" LIVE STREAMING	AOWSAf8Tx18	https://www.you	2025-01-06T05:39:21Z	@muhajarohnizam419	Towel seneng liat sty di pecat		
1266	ðŸŽŸ" LIVE STREAMING	AOWSAf8Tx18	https://www.you	2025-01-06T05:39:21Z	@budimanculiktv8139	Towel out dari indone		
1267	ðŸŽŸ" LIVE STREAMING	AOWSAf8Tx18	https://www.you	2025-01-06T05:39:21Z	@AnahSunarsih	TOWEL KETAWA LEB		
1268								

HASIL ANALISIS SENTIMEN

Berdasarkan kasus yang dianalisis dan tujuan yang dimaksudkan maka dibuatlah sebuah *Workflow* yang digunakan dan dirancang pada Orange Data Mining seperti yang terlihat pada gambar dibawah ini.



Widget Digunakan Pada Orange Data Mining

Pada Workflow di atas, berikut penjelasan tiap bagian beserta widget yang digunakan:

a. Corpus

- 1) Widget ini digunakan untuk mengimpor, mengelola, dan menyimpan data berbasis teks (seperti dokumen, komentar, atau tweet).
- 2) Corpus menjadi sumber data utama untuk proses analisis teks berikutnya.

b. Preprocess Text

- 1) Membersihkan dan mempersiapkan data teks untuk analisis lebih lanjut.
- 2) Termasuk langkah-langkah seperti *tokenization*, menghapus *stop words*, normalisasi teks (*lowercasing*), dan penghapusan simbol.

c. Sentiment Analysis

- 1) Menganalisis teks dan mengklasifikasikan sentimen menjadi kategori seperti positif, negatif, atau netral.
 - 2) Menggunakan model analisis sentimen bawaan atau custom.
- d. Word Cloud
- 1) Memvisualisasikan kata-kata yang paling sering muncul dalam dataset teks dalam bentuk awan kata.
 - 2) Memberikan gambaran awal tentang topik atau istilah yang sering dibahas.
- e. Distribution
- 1) Menampilkan distribusi data dalam bentuk grafik (seperti histogram, pie chart, atau bar chart).
 - 2) Dalam konteks sentimen, widget ini menunjukkan proporsi data berdasarkan kategori sentimen.
- f. Tweet Profiler
- 1) Memberikan analisis lebih rinci tentang setiap teks dalam dataset, termasuk kategori sentimen, skor sentimen, dan fitur tambahan lainnya.
 - 2) Berguna untuk memahami profil teks secara individu.

Proses Analisis

Berdasarkan data yang sudah diperoleh pada proses crawling data yang diambil dari komentar video youtube dengan *keyword* “Copot Shin Tae-Yong” & “Copot STY”, penulis ingin melakukan sebuah analisis dengan proses sebagai berikut :

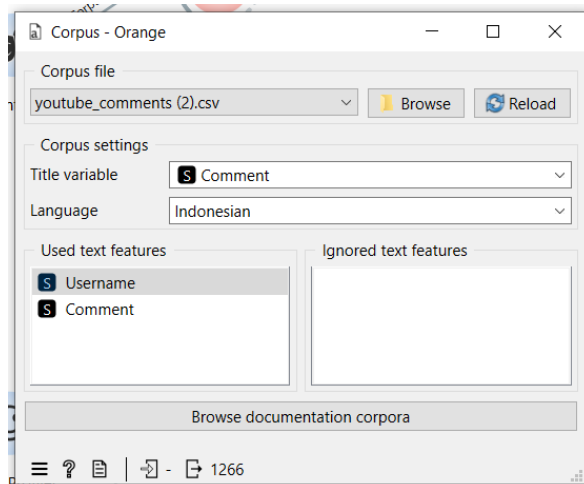
- a. Mengumpulkan semua data hasil scrawling dari semua komentar youtube dengan query pencarian “Copot Shin Tae-Yong” & “Copot STY” dimana sebelumnya data sudah disortir dan dipastikan sesuai dengan query yang diminta.
- b. Melakukan sebuah filterisasi dengan mempertimbangkan kata-kata, karakter, singkatan, bahkan tanda baca yang tidak ada hubungannya dengan data yang akan dilakukan analisis.
- c. Melakukan sebuah proses Sentiment Analysis dari data yang telah dihasilkan dari filterisasi, dengan memberikan pertimbangan kata positif, dan negatif tambahan sesuai kebutuhan untuk menghindari miss asumsi dari data tersebut.
- d. Melakukan sebuah proses *Emotion Profiler (Tweet Profiler)* dari data yang telah dihasilkan menghasilkan sebuah profil seperti Anger, Joy, Disgust, dll.

- e. Menghasilkan sebuah diagram untuk melihat perbandingan dari masing-masing sentiment dan Emotion Profiler yang dianalisis.
- f. Melakukan sebuah pembagian terhadap sentimen analisis sebagai berikut :
 - 1) Nilai sentiment < 0 \rightarrow “Negatif”
 - 2) Nilai sentiment $= 0$ \rightarrow “Netral”
 - 3) Nilai sentiment > 0 \rightarrow “Positif”
- g. Membuat sebuah word cloud dari pembagian nilai sentiment untuk melihat apa saja kata yang sering muncul sebagai kata Negatif, Netral dan Positif.

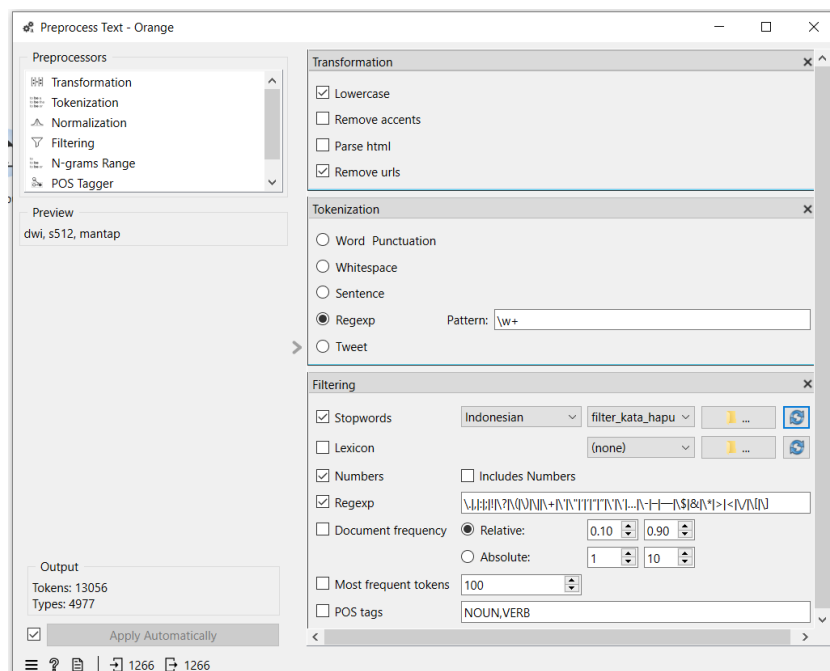
Workflow Orange Data Mining

Dari proses yang sudah didefinisikan maka penulis membuat sebuah *workflow* pada *orange data mining*. Berikut urutan pengerjaan dari setiap komponen *workflow*-nya:

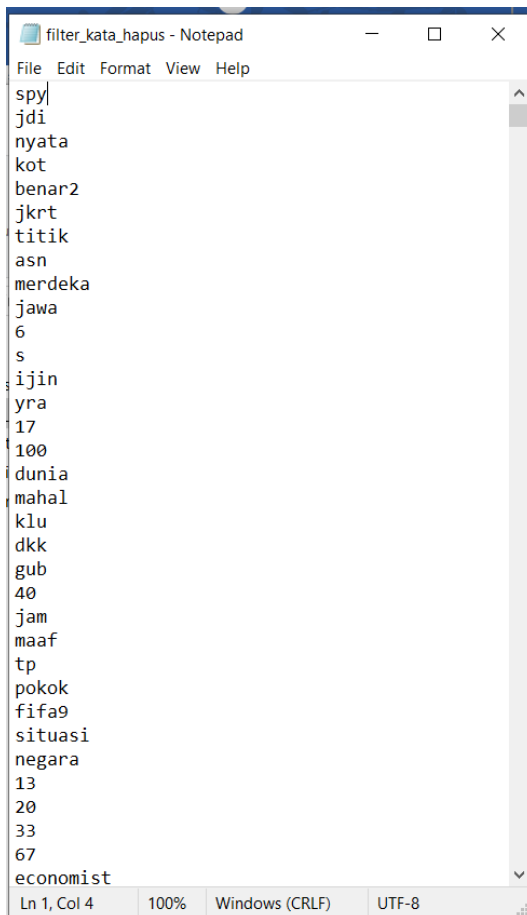
1) Import data pada widget Corpus



2) Melakukan Filterisasi dari dataset yang sudah di import dengan widget Preprocess Text.

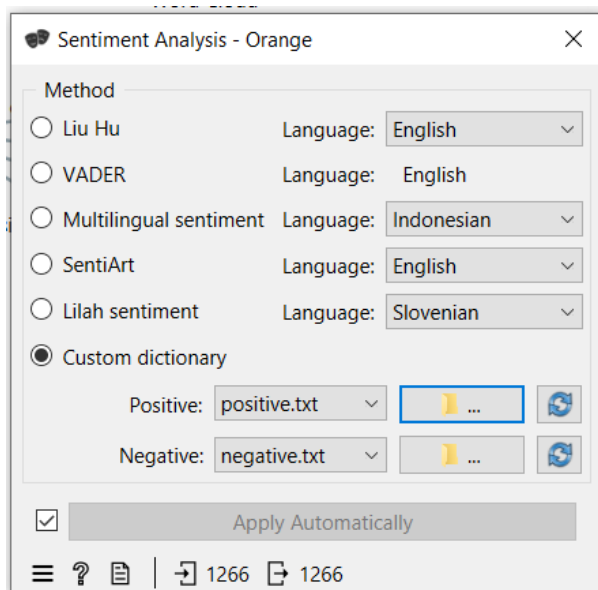


3) Menambahkan filter stopwords custom pada Widget Preprocess Text (1163 kata).

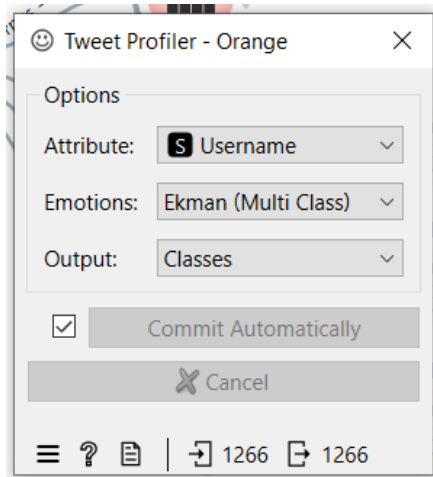


4) Melakukan Anlisa Sentimen menggunakan widget Sentiment Analysis.

Selanjutnya yaitu menambahkan custom dictionary untuk identifikasi kata positif dan negatif. Referensi <https://github.com/masdevid/ID-OpinionWords>

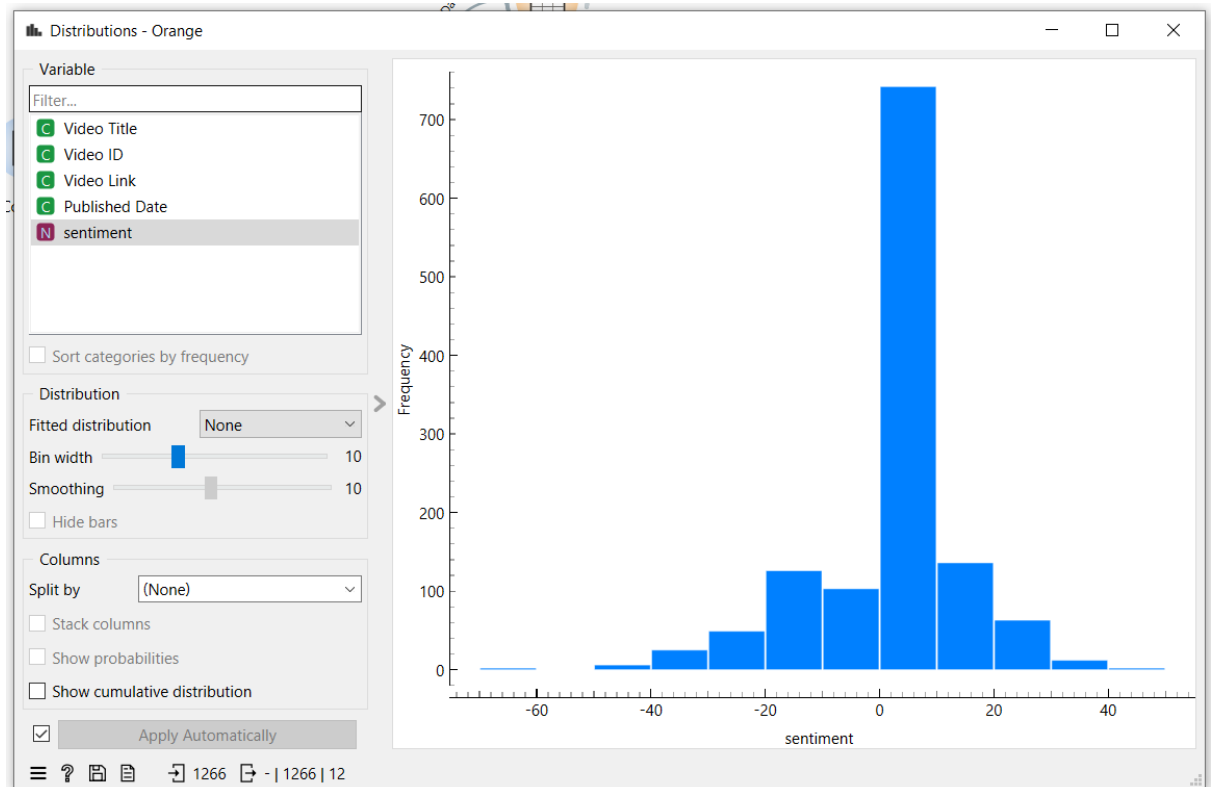


5) Melakukan Analisa Emotion Profiler menggunakan Widget Tweet Profiler

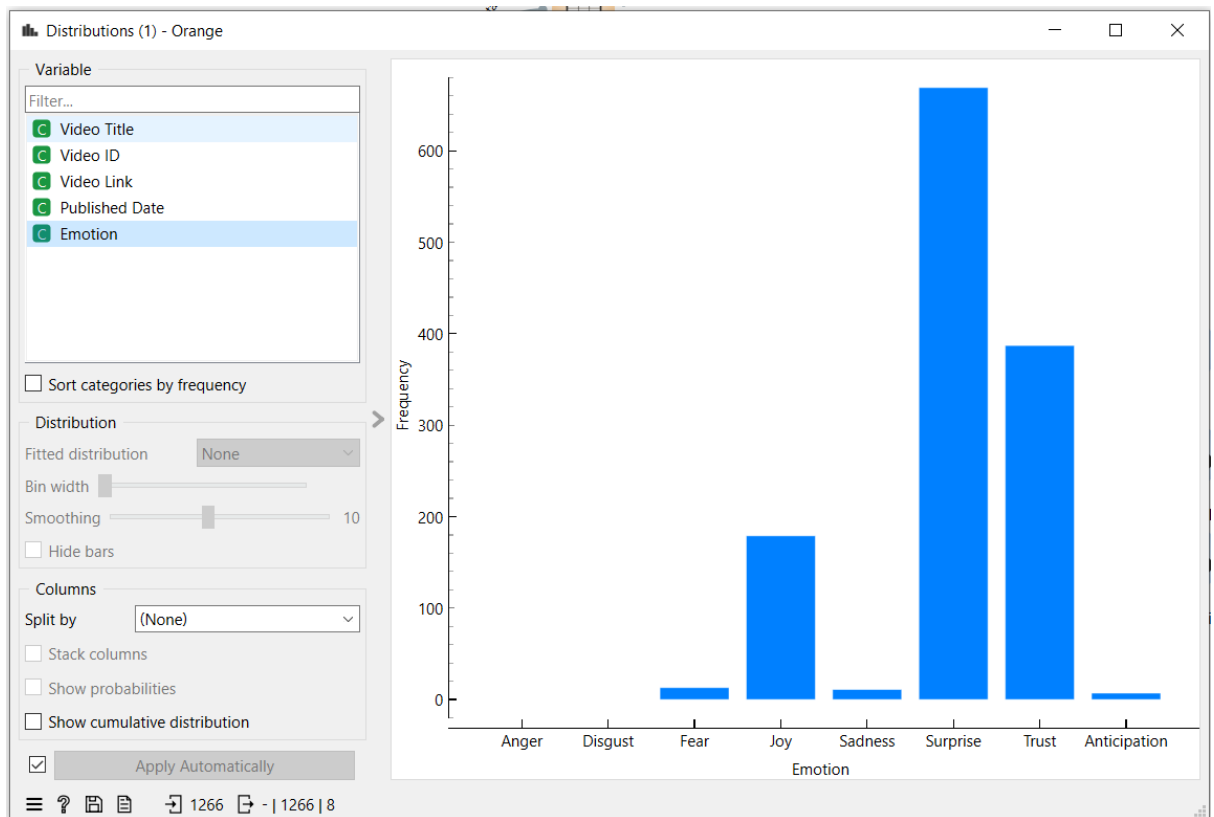


- 6) Melihat detail Distribution Sentiment Analysis dan Tweet Profiler dan menampilkan hasil analisisnya menggunakan Widget Distribution dan Word Cloud.

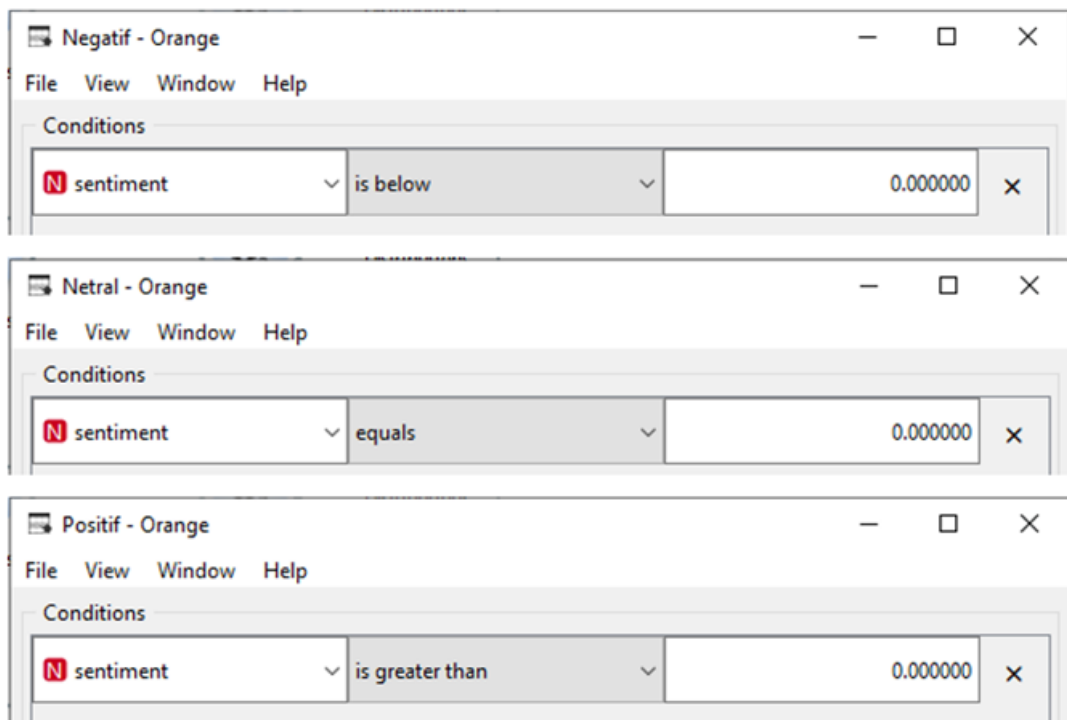
Sentimen Analisis :



Tweet Profiler/Emotion :

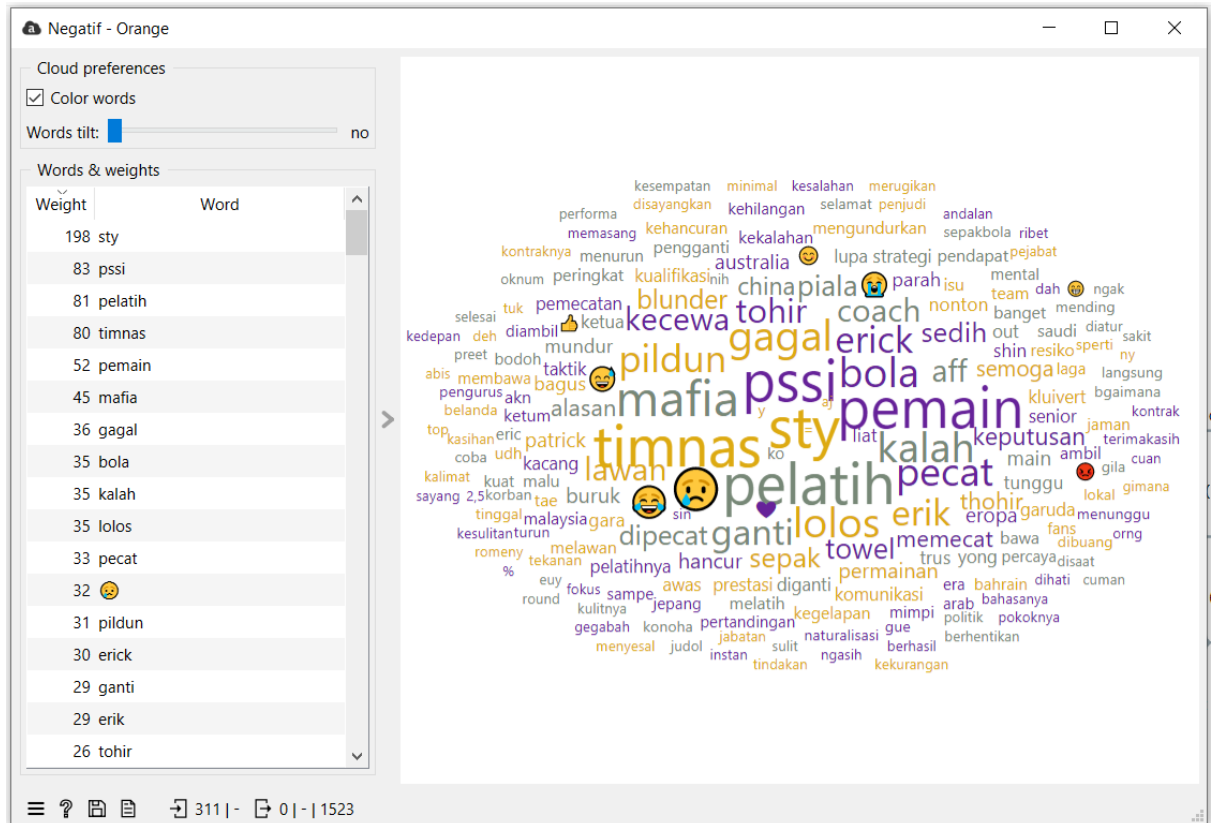


- 7) Melakukan pembagian data berdasarkan Sentimen Negatif, Netral dan Positif dan ditampilkan menggunakan word count dengan menggunakan kualifikasi pembagiann sebagai berikut:

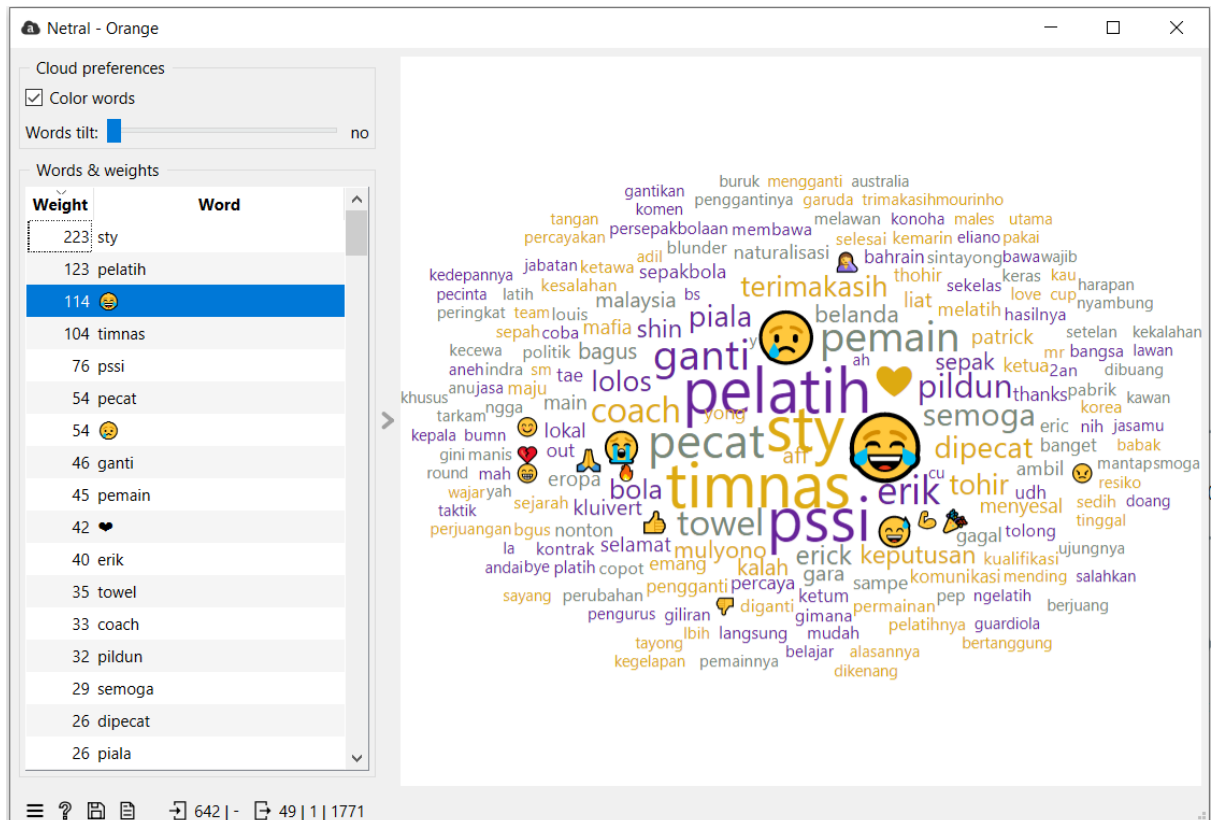


Berikut adalah hasil Sentimen Analisis pada word cloud:

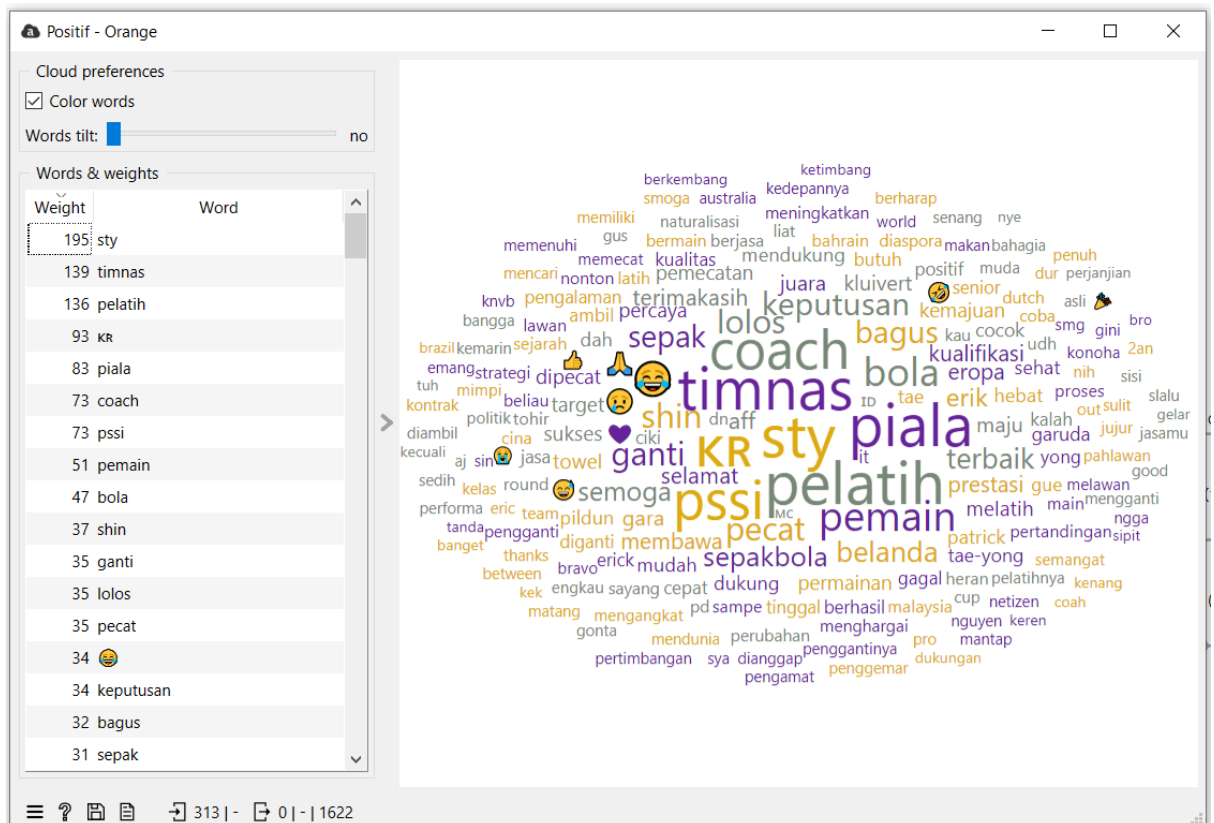
8) Sentimen Negatif



9) Sentimen Netral



10) Sentimen Positif



KESIMPULAN

Berdasarkan kedua gambar nilai *Distribution* pada Analisis Sentimen dan *Tweet Profiler/Emotion* yang ada maka dapat disimpulkan sebagai Berikut :

1. Sentimen Analisis

- a. Diagram pada sentimen analisis menunjukkan bahwa sebagian besar sentimen yang terkait dengan pemberitaan mengenai pemecatan Shin Tae-yong berada di kisaran netral, dengan puncak distribusi berada di nilai sentimen sekitar nol.
- b. Hal ini menunjukkan bahwa publik cenderung memberikan opini yang bersifat faktual atau tanpa melibatkan emosi yang dominan dalam diskusi terkait. Namun, terdapat beberapa opini dengan sentimen negatif (di bawah nol) dan positif (di atas nol), meskipun jumlahnya relatif lebih kecil dibandingkan sentimen netral.

2. Tweet Profiler/Emotion

- a. Grafik distribusi emosi menunjukkan bahwa **surprise (kejutan)** menjadi emosi dominan, diikuti oleh **trust (kepercayaan)**. Hal ini mengindikasikan bahwa banyak pihak yang terkejut dengan keputusan pemecatan Shin Tae-yong.
- b. Di sisi lain, tingkat kepercayaan yang signifikan menunjukkan bahwa sebagian masyarakat masih percaya terhadap alasan atau keputusan yang diambil oleh PSSI terkait pemecatan ini.
- c. Emosi lain seperti **fear (ketakutan)** juga muncul dengan frekuensi yang cukup besar, yang mungkin mencerminkan kekhawatiran akan dampak keputusan tersebut terhadap masa depan TimNas Indonesia.
- d. Emosi negatif seperti **anger (kemarahan)** dan **disgust (kejiikan)** muncul dalam jumlah yang sangat kecil, menunjukkan bahwa respon negatif emosional terhadap berita ini tidak terlalu signifikan.

Ringkasan

Dari analisis sentimen dan emosi, dapat disimpulkan bahwa pemberitaan terkait pemecatan Shin Tae-yong sebagai pelatih Tim Nasional Sepak Bola Indonesia memicu diskusi publik yang didominasi oleh sentimen netral, dengan emosi kejutan sebagai respons utama.

Masyarakat tampaknya terpecah antara rasa kepercayaan terhadap otoritas yang mengambil keputusan dan kekhawatiran atas dampak yang mungkin timbul. Reaksi negatif seperti kemarahan dan kebencian relatif kecil, yang menunjukkan bahwa keputusan ini diterima dengan beragam opini, namun tidak mengarah pada respons emosional ekstrem.

Kesimpulan ini mengindikasikan bahwa isu pemecatan ini merupakan peristiwa yang mencuri perhatian publik, namun tidak secara drastis memicu polemik atau kontroversi yang meluas.