**LAPORAN CRAWLING, CLEANING, SENTIMENT ANALYS KOMENTAR YOUTUBE**

Devina Dewi Faustina Basam (20220801250)

Eneng Mila Lestari Wijayadi (20220801264)

1. **Deskripsi Kasus**

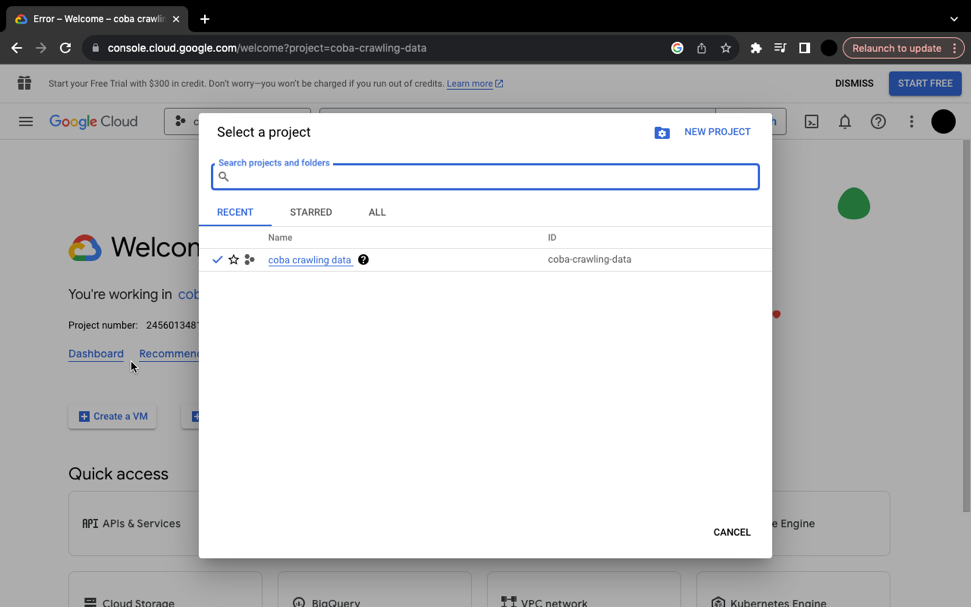
Membuat program Engine Crawling untuk menangkap komentar pengguna suatu produk pada youtube menggunakan Bahasa Phyton. Dan nantinya data yang didapat berupa kumpulan komentar di suatu video youtube, kapan komentar tersebut dibuat, nama author yang menulis komentar tersebut, dan berapa banyak orang yang menyukai komentar tersebut. Data Komentar tersebut selanjutnya akan di cleaning menggunakan Rapid Miner. Lalu datanya masuk ke tahap terakhir untuk dilakukan analisis sentimen pada data komentar yang telah di cleaning menggunakan Bahasa Phyton untuk menemukan sentimen mayoritas di koemntar suatu youtube.

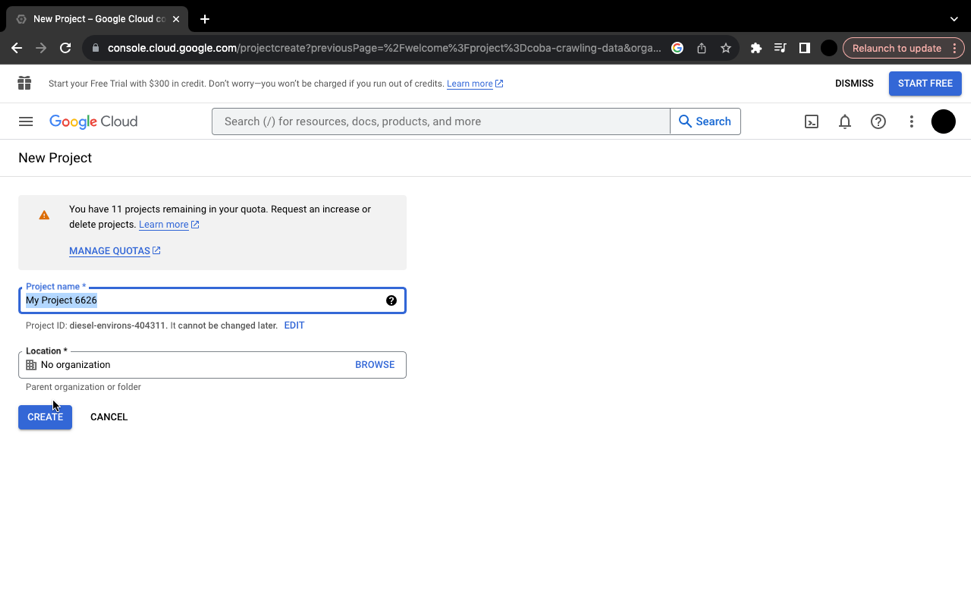
1. **Data Crawling Menggunakan Google Colab**

Dalam melakukan data crawling kami menggunakan Google Colab untuk mengimplementasikan Bahasa Pemrograman Phyton dan menggunakan Google Developer Console agar mendapatkan API key untuk dapat mengakses youtube.

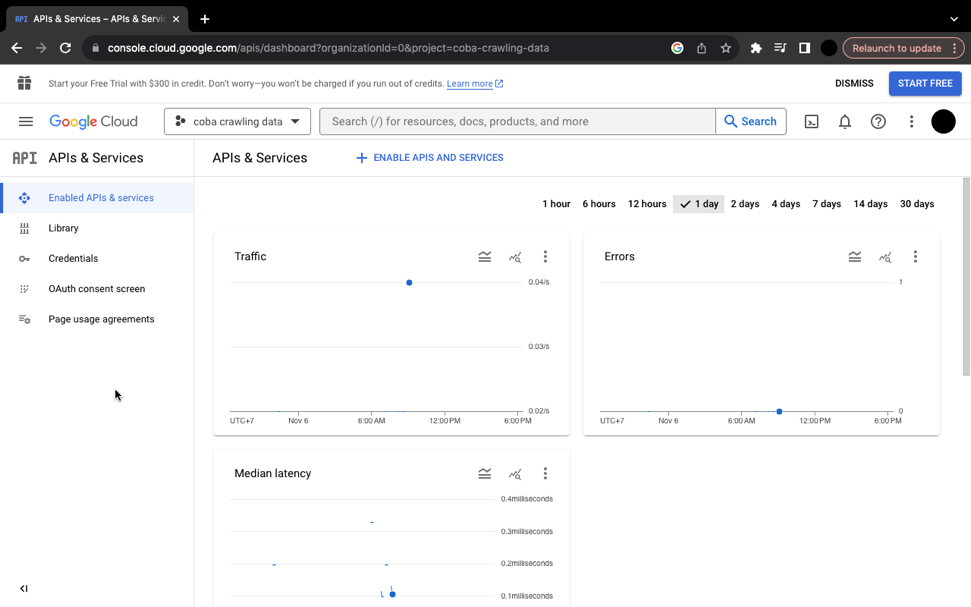
Untuk mendapatkan API Key di Developer console dengan cara sebagai berikut :

1. Membuka <https://console.cloud.google.com> dan membuat proyek baru

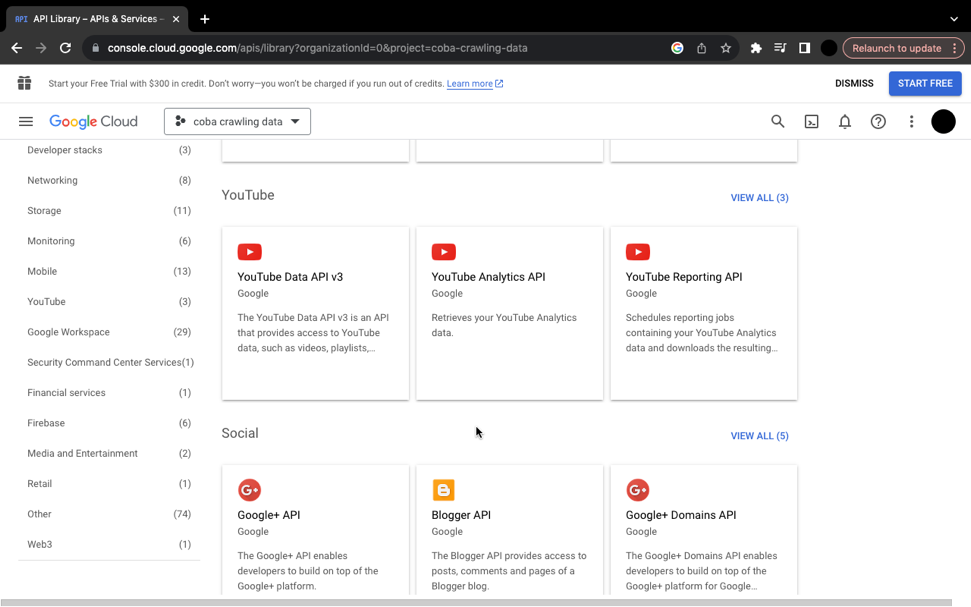


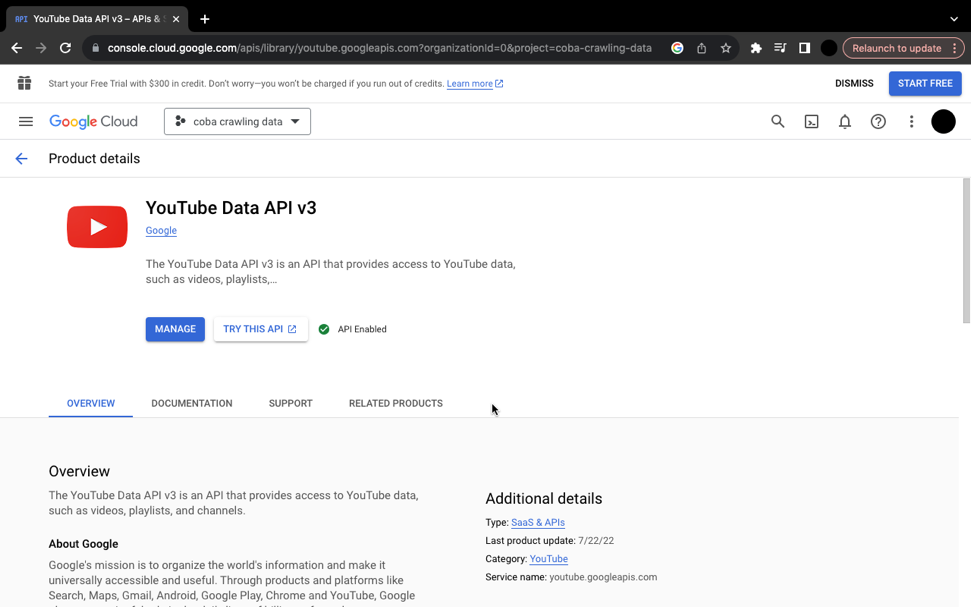


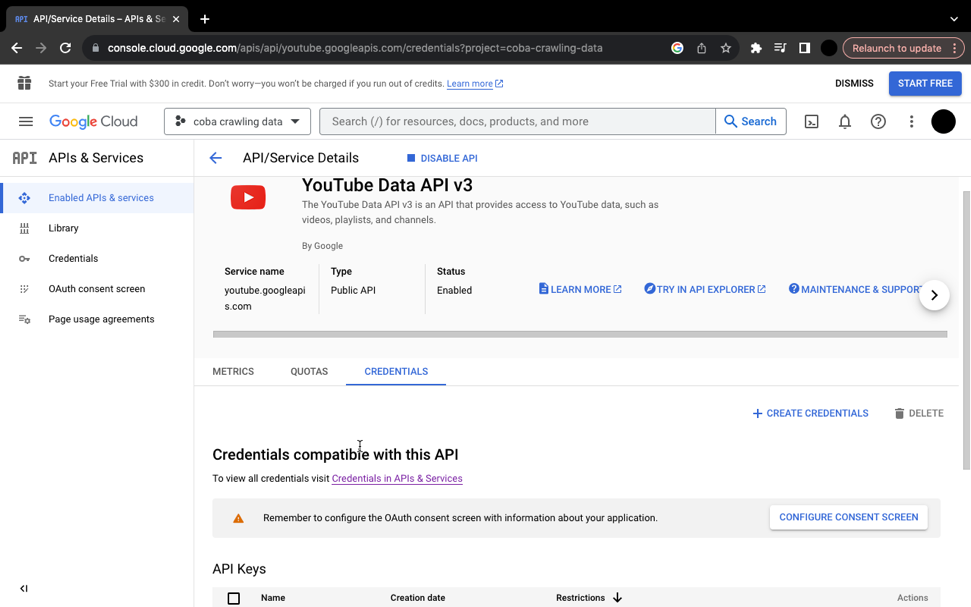
1. Menekan tiga titik garis disebelah kiri pojok dan memilih menu APIs & Services dan ke bagian ENABLE APS AND SERVICES



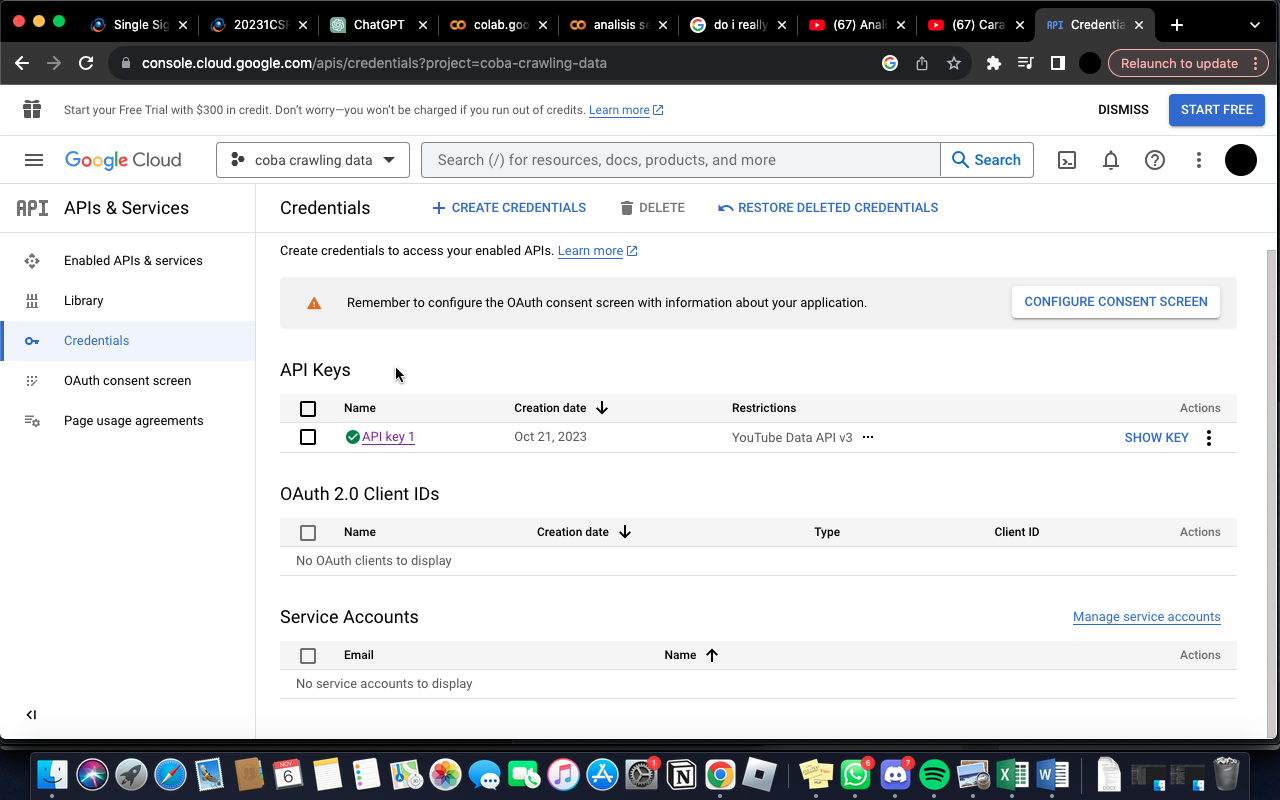
1. Di bagian ENABLE APS AND SERVICES, scroll kebawah sampai menemukan bagian youtube dan pilih YouTube Data API v3



1. Setelahnya tekan MANAGE untuk mengaktifkan YouTube Data API v3
2. Pilih CREDENTIALS lalu tekan CREATE CREDENTIALS

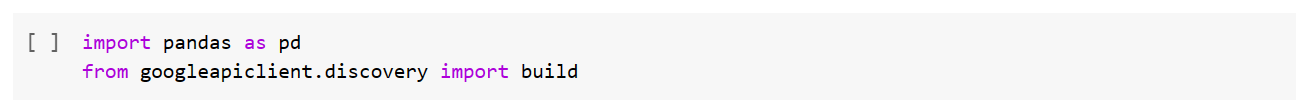


1. Setelah membuat credentials kita akan mendapatkan API key seperti dibawah ini dan untuk menunjukkan API Key nya tekan SHOW KEY



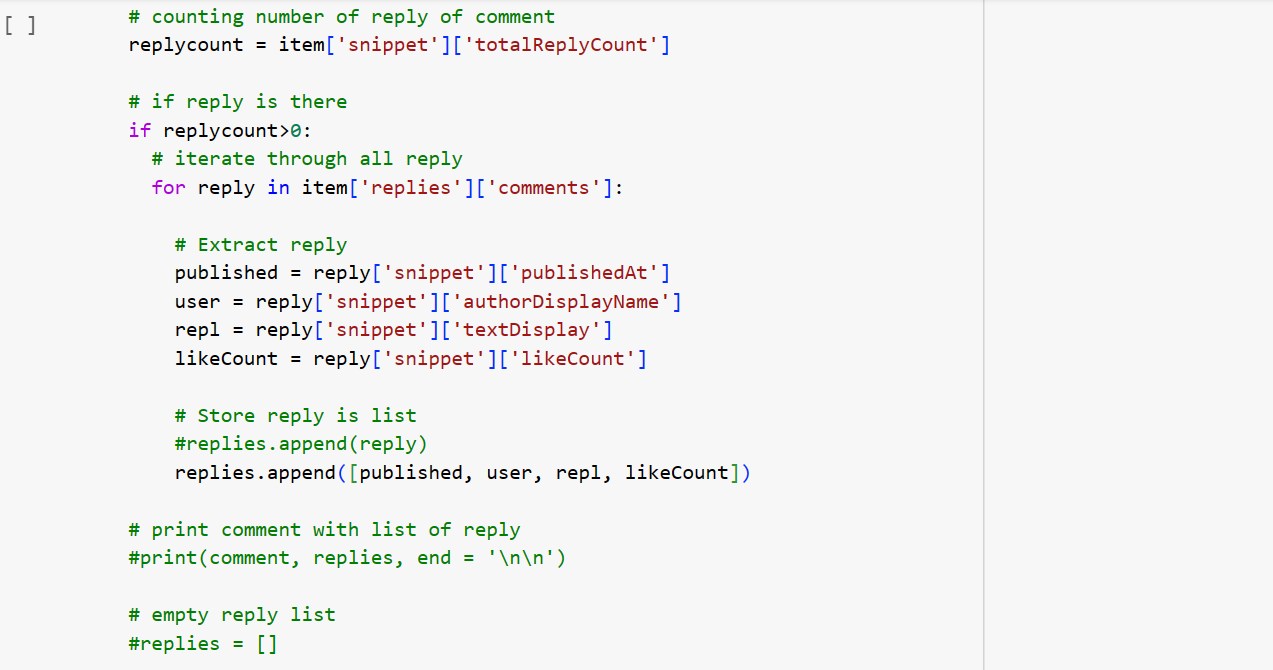
Implementasi Coding Phyton untuk crawling komentar dari video di youtube yang kami gunakan adalah sebagai berikut :

1. Import library menggunakan Pandas & Google Api Client.

****

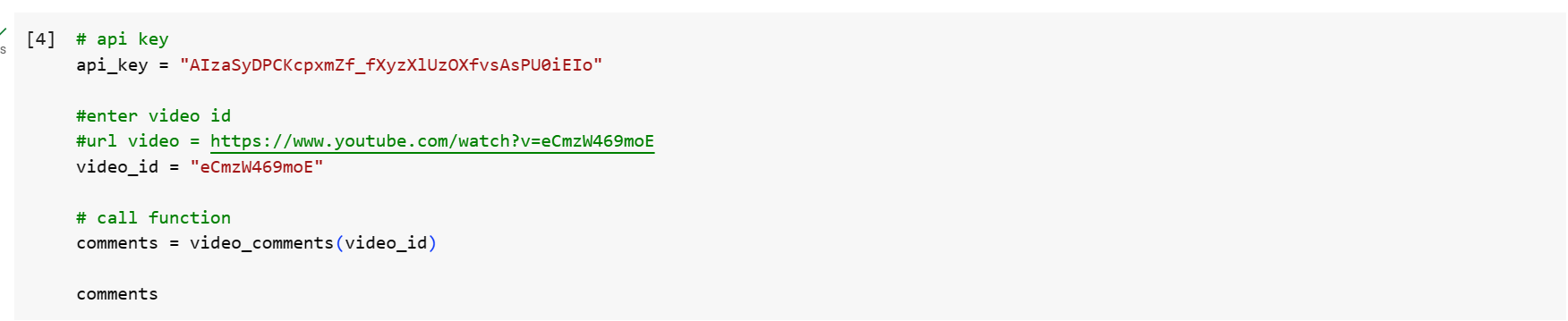
1. Fungsi untuk Crawling komentar youtube dengan cara memanggil library dari youtube video produk yang akan diambil datanya selanjutnya fungsi ini akan return data komentar dari video youtube tersebut.

****

****

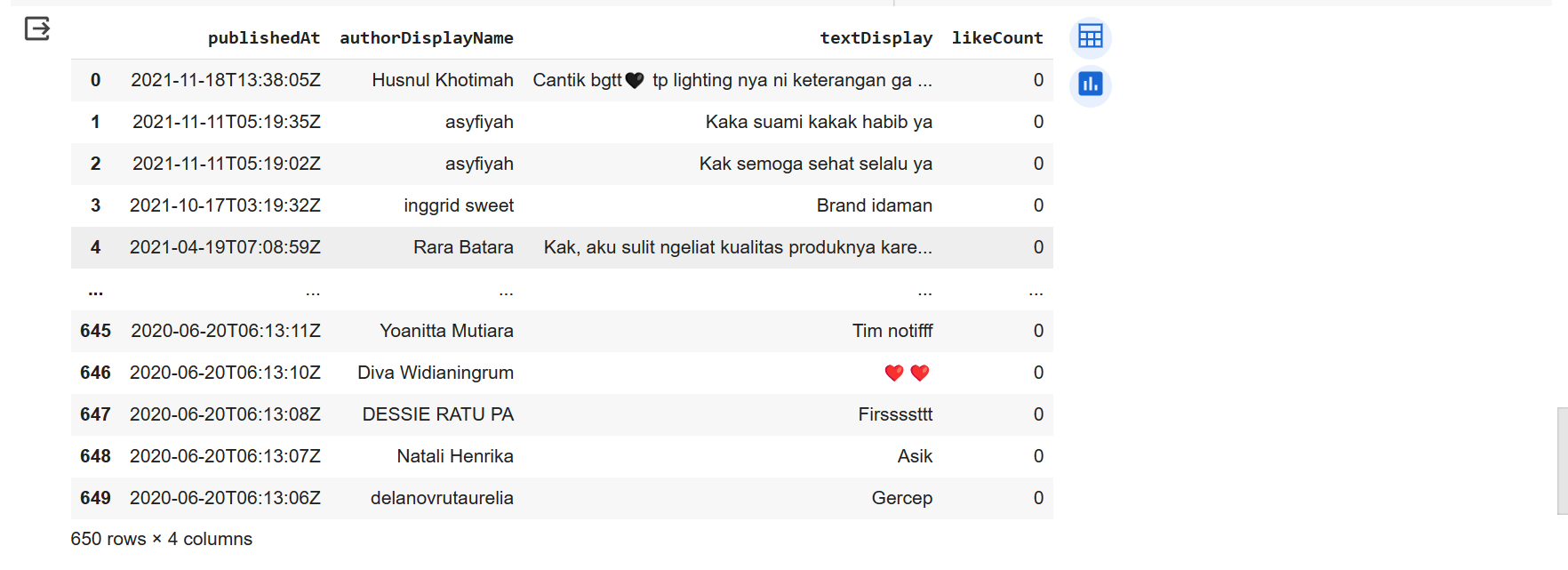
****

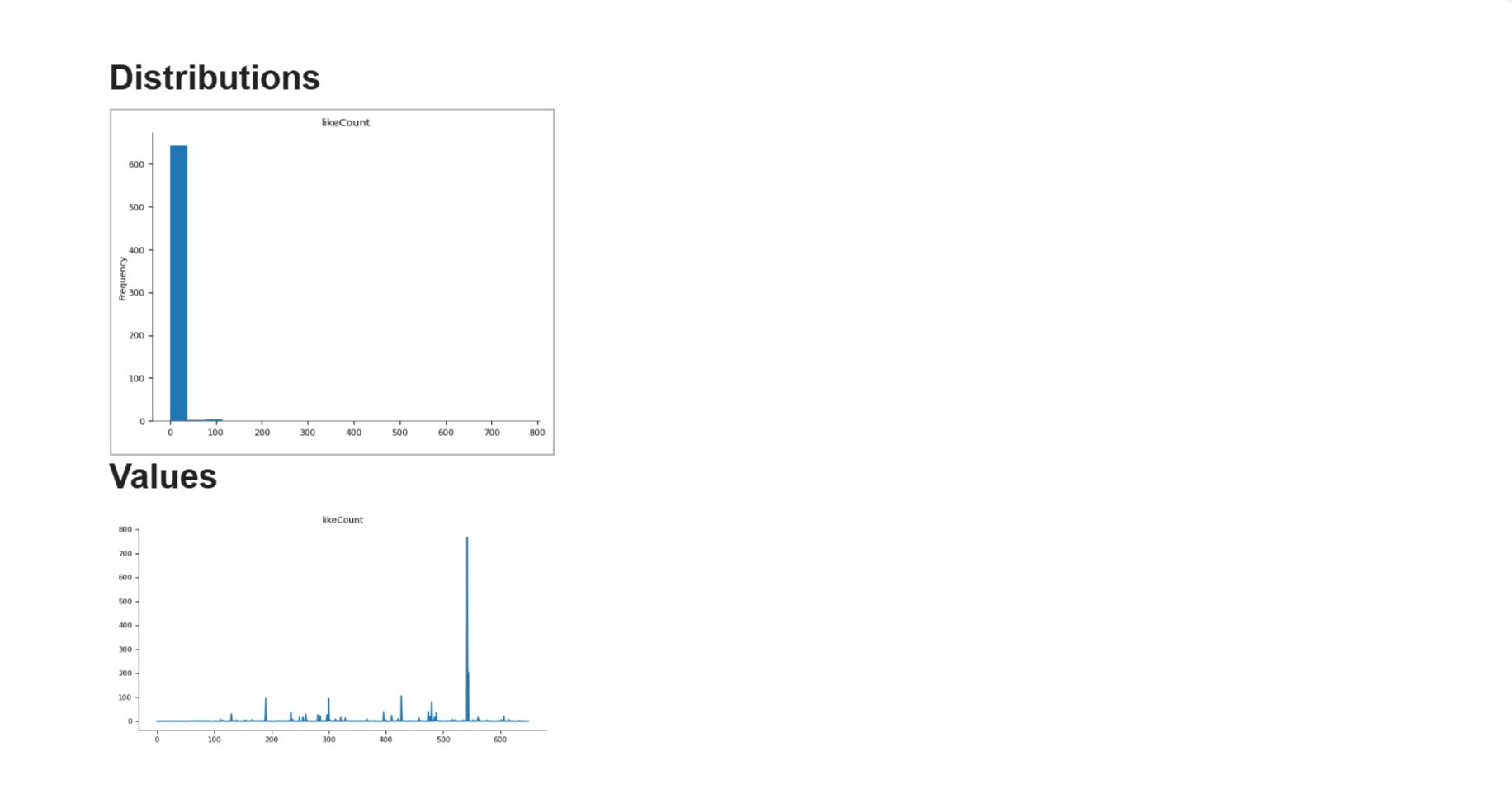
1. Menjalankan Fungsi Crawling komentar menggunakan API key dan Link Youtube video yang akan diambil data komentarnya.

****

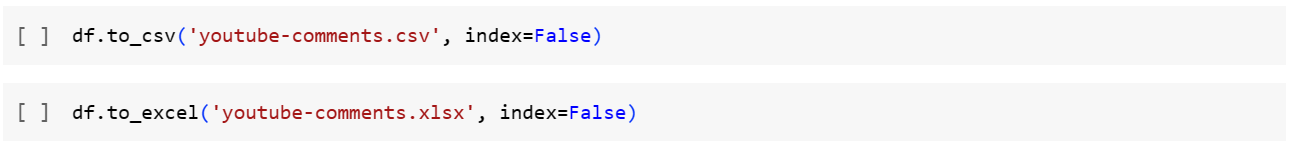
****

1. Mengubah data Crawling komentar kedalam DataFrame menggunakan Pandas.

****

****

1. Menyimpan Hasil Crawling ke file CSV & Excel

****

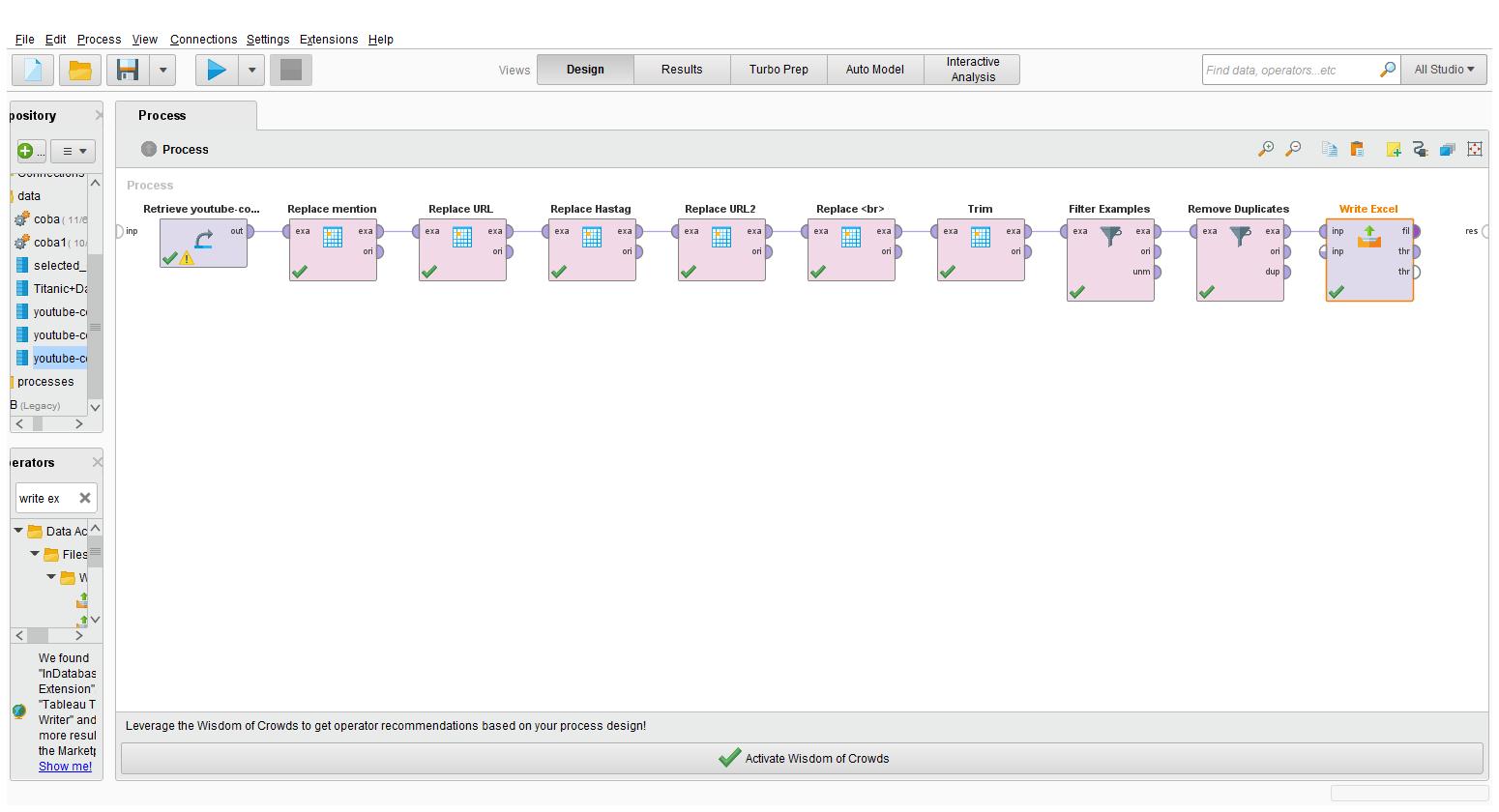
1. Implementasi Codingan diatas dapat diakses pada link berikut;

<https://colab.research.google.com/drive/1sbz42Gt1ypAHdBRWxwMLA7LrSyMVRCxw?usp=sharing>

1. **Cleaning Data Menggunakan Rapid Miner**

Cleaning data ini dilakukan untuk membersihkan data crawling komentar dari karakter yang tidak diperlukan dan mengkonversi teks kedalam format yang sesuai untuk analisis.

Dalam Cleaning Data Komentar Youtube kami menggunakan tools sebagai berikut:

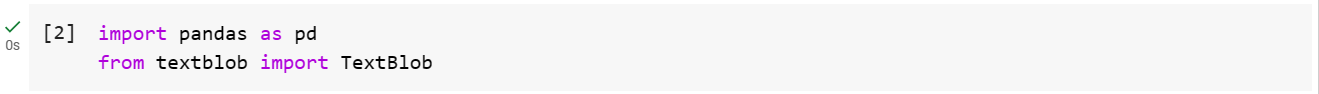


1. **Replace Mention** : Digunakan untuk menghilangkan mention. Dengan mengetik @.\*? di bagian replace what. Atribut filter type “single” dan attribute “textDisplay”.
2. **Replace URL 1** : Menghilangkan link yang tertera di depan dan di tengah pada komentar. Dengan mengetik <a href.\*? di bagian replace what. Atribut filter type “single” dan attribute “textDisplay”.
3. **Replace URL 2**  : Menghilangkan Link yang tertera didepan dan tengah pada komentar. Dengan mengetik <a href.\* di bagian replace what. Atribut filter type “single” dan attribute “textDisplay”.
4. **Replace <br>** : Menghilangkan <br>. Dengan mengetik <br> di bagian replace what. Atribut filter type “single” dan attribute “textDisplay”.
5. **Replace Hastag** : Menghilangkan Hastag dalam data. Dengan mengetik #.\*? di bagian replace what. Atribut filter type “single” dan attribute “textDisplay”.
6. **Trim** : Menghilangkan Spasi yang terlalu banyak.
7. **Filter Examples** : Untuk menampilkan data yang memiliki isian dan tidak menampilkan data yang tidak memiliki isian dengan menggunakan custom filter “is not missing” dan value “textDisplay”.
8. **Remove Duplicate** : Menghilangkan data yang memiliki duplicate. Atribut filter type “single” dan attribute “textDisplay”.
9. **Write Excel** : Untuk Mengekspor hasil Pemrosesan data ke File Excel dan mendapatkan output file “after cleaning.xlsx”
10. **Analisis Sentiment Menggunakan Google Colab**

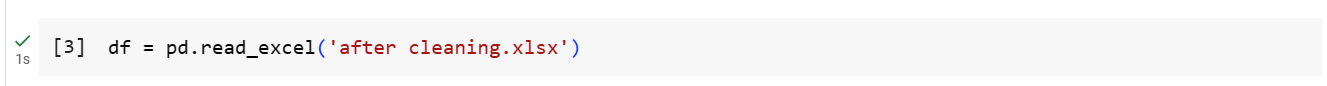
Analisis sentimen adalah proses untuk menilai atau mengevaluasi sentimen, opini, atau perasaan yang terkandung dalam teks, ulasan, atau data lainnya. Ini sering digunakan dalam berbagai konteks, termasuk bisnis, media sosial, survei pelanggan, dan analisis pendapat. Dalam data comment youtube ini kami menggunakan analisis sentiment untuk menampilkan hasil positif, negative, dan netral.

Implementasi Coding Phyton yang kami gunakan adalah sebagai berikut;

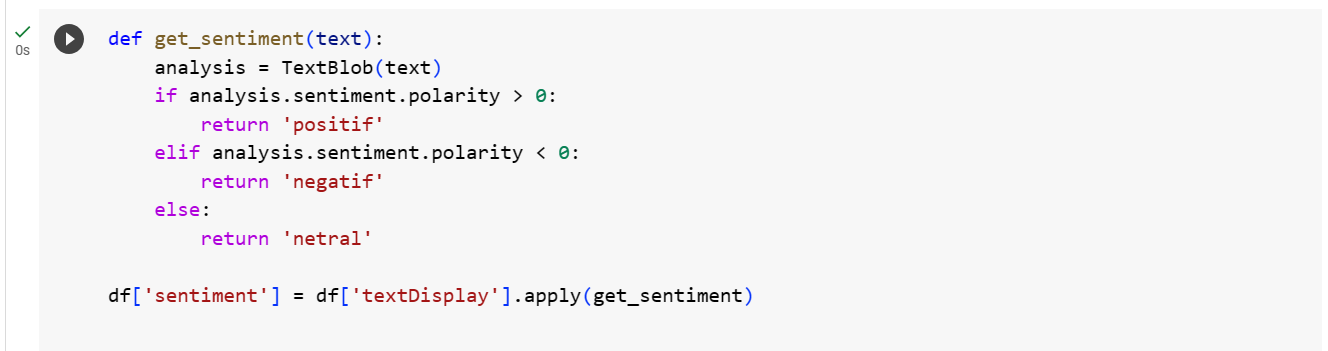
1. Import library menggunakan Pandas dan TextBlob.



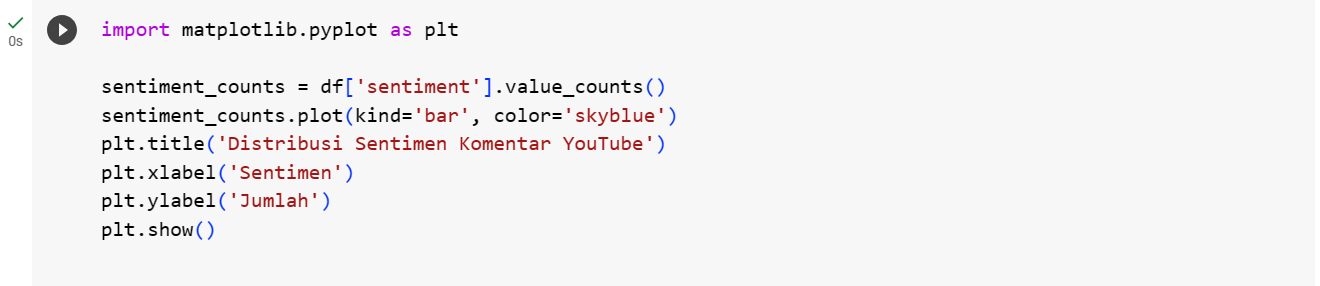
1. Memasukan data yang dibaca dari data excel “after cleaning.xlsx” ke dalam data frame.

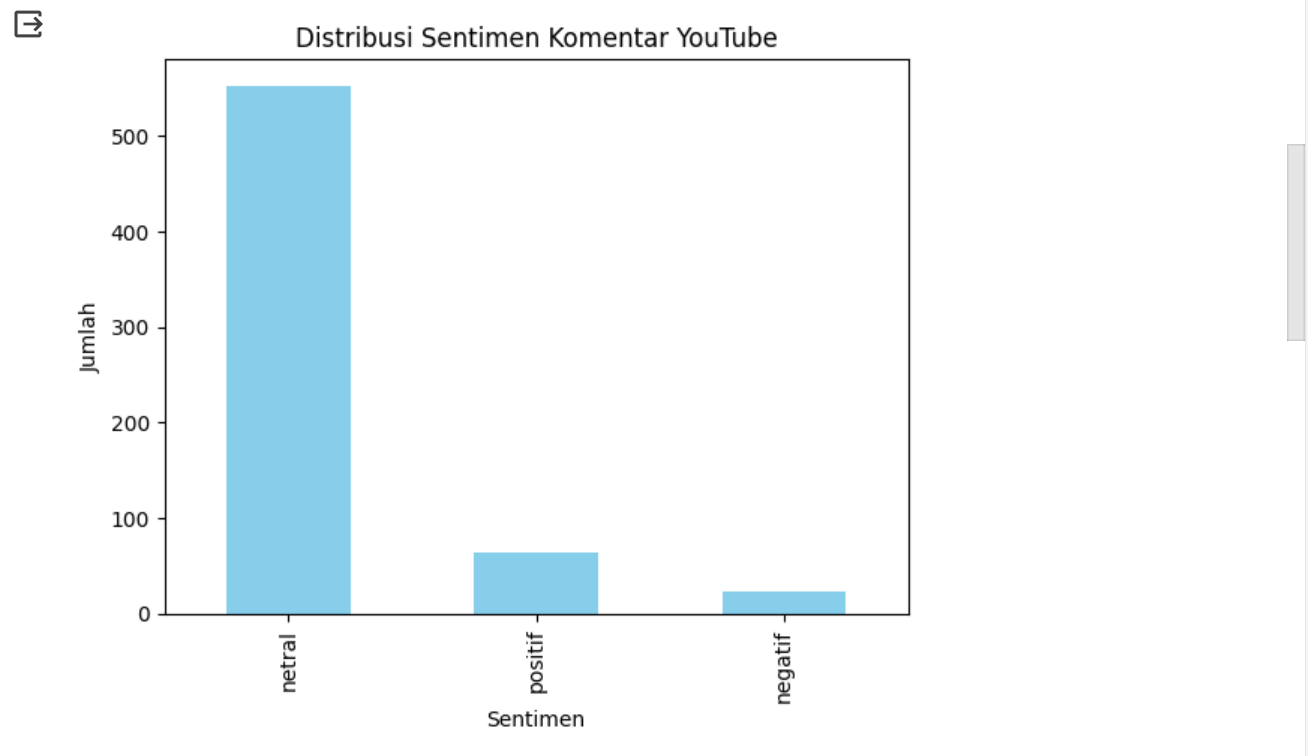


1. Implementasi untuk melakukan analisis sentiment.



1. Memunculkan diagram batang.





1. Mengubah menjadi tabel.





1. Menyimpan Hasil Crawling ke file CSV & Excel

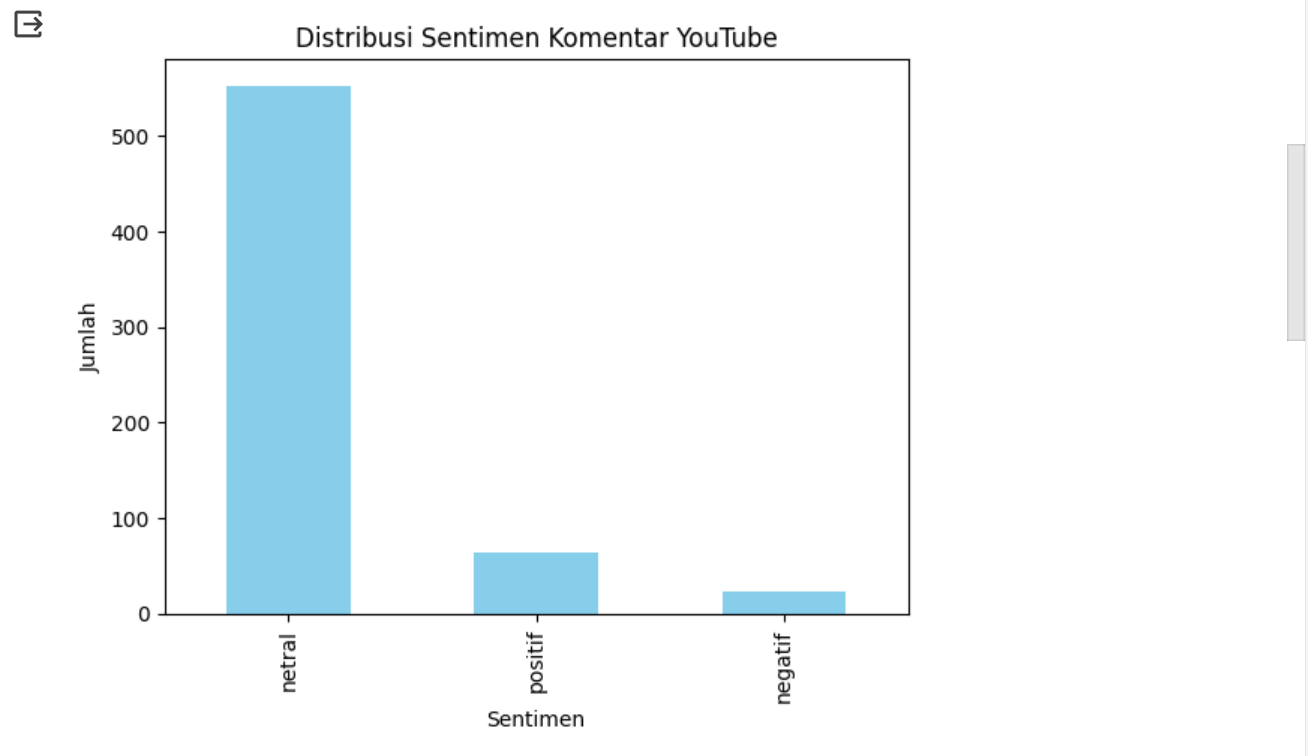
****

1. Implementasi Codingan diatas dapat diakses pada link berikut;

<https://colab.research.google.com/drive/1GlpH5HIHqY0BhSi6QOTQAgt9h3AzMHfx?usp=sharing>

1. **Visualisasi**

Visulisasi yang kami berikan berupa diagram batang distribusi dari analisis sentiment :



Dengan banyak komentar dengan sentiment netral sebanyak 553 komentar, positif 64 komentar dan negatif 23 komentar.

1. **Intepretasi**

Berdasarkan hasil dari crawling komentar dari video youtube tersebut dan dilakukannya analisis sentiment didapatkan bahwa mayoritas komentar tersebut adalah komentar bersentimen netral.

File yang didapatkan selama proses proses diatas :

* Data hasil crawling = youtube-comments.xlsx
* Data hasil crawling yang sudah di cleaning di rapidminer = after cleaning.xlsx
* Data hasil analisis sentiment = hasil\_analisis\_sentimen.xlsx