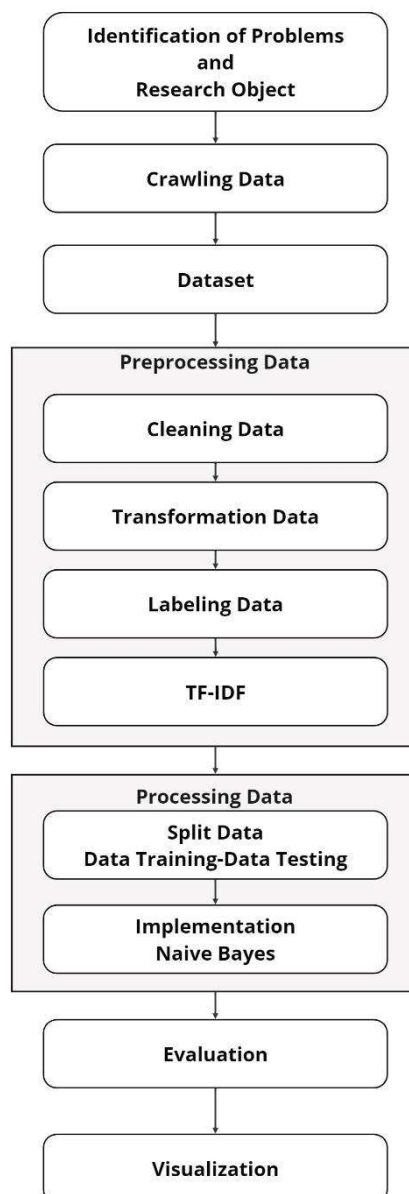


BAB III

METODE PENELITIAN

3.1. Desain Penelitian

Tahapan dalam penelitian ini akan melalui beberapa proses yang akan dijelaskan secara rinci dan digambarkan pada **gambar 3.1**.



Gambar 3. 1 Desain Penelitian
(Sumber: Peneliti 2024)

Keterangan dari Desain Penelitian:

1. Identifikasi Masalah dan Objek Penelitian

Korlantas POLRI telah mengusulkan penggunaan *electronic traffic law enforcement (ETLE)* yang bertujuan untuk mendukung keamanan, keselamatan, dan ketertiban dalam pelaksanaan lalu lintas. Keputusan ini telah menimbulkan berbagai reaksi di kalangan masyarakat. Tujuan penelitian ini adalah untuk menganalisis sentimen masyarakat terhadap pro dan kontra terkait penerapan ETLE. Objek penelitian ini adalah opini masyarakat mengenai isu tersebut dalam *platform* media sosial Twitter.

2. Crawling Data

Dalam proses ini sumber data yang diambil merupakan data *tweet* yang diambil secara langsung dari sosial media Twitter. Proses *crawling* ini dilakukan dengan *platform* Google Colab menggunakan bahasa pemrograman Python.

3. Dataset

Dataset yang digunakan dalam penelitian ini berbentuk format *comma separated values (CSV)* yang didapatkan dari *database* Twitter dengan metode *crawling*. Data yang dihasilkan merupakan data acak tanpa memperhitungkan pemilik akun baik secara asli, milik media serta akun yang terindikasi palsu.

4. **Cleaning Data**

Sebelum melakukan proses klasifikasi, penting untuk melakukan *preprocessing* data komentar Twitter agar mengurangi dimensi model ruang vektor. *Preprocessing* data dibagi ke dalam dua tahap. Tahap pertama *cleaning data* bertujuan untuk membersihkan data dari simbol-simbol, *link web*, *mention*, dan *retweet*.

5. **Transformasi Data**

Tahap kedua pada *preprocessing* data adalah transformasi data, tahap ini akan dilakukan perubahan teks menjadi *lower case (case folding)*, mengubah setiap kalimat menjadi kata-kata (*tokenizing*), membuang kata-kata yang tidak bernilai (*stopword removal*), dan membuang kata-kata yang terlalu pendek atau panjang (*filter by length*). (Ramlan et al., 2023).

6. **Labeling Data**

Setelah dataset dibersihkan selanjutnya data diberi label yaitu positif dan negatif sesuai dengan kategori atau klasifikasi yang relevan dengan tujuan penelitian. Proses ini dilakukan secara manual berdasarkan dari analisis peneliti.

7. **TF-IDF**

Pemberian bobot pada setiap kata merupakan teknik umum dalam pencarian informasi. Teknik ini dikenal sebagai TF-IDF (*Term Frequency–Inverse Document Frequency*). Metode ini menilai pentingnya suatu istilah atau kata dalam sebuah dokumen, dan dianggap sebagai pendekatan yang sederhana, akurat, dan efisien. Berikut adalah rumus TF-IDF:

$$W(t.d) = Wtf(t.d) \times idft \quad \text{Rumus 3.1 TF-IDF}$$

Keterangan:

$W(t.d)$: Bobot TF-IDF

$Wtf(t.d)$: Bobot kata dalam setiap dokumen

$idft$: Bobot Inverse Document Frequency dari nilai $(\log(N/df))$

N : Jumlah seluruh dokumen

Df : Jumlah dokumen yang mengandung term

8. Split Data

Pemisahan data melibatkan penentuan apakah data akan digunakan sebagai data pelatihan (*training*) atau data pengujian (*testing*). Sebagian data yang telah dilabeli akan digunakan sebagai data pelatihan, yang akan dimanfaatkan oleh model *machine learning* untuk mempelajari pola dari fitur-fitur yang ada. Sisanya dari data yang telah diberi label akan dijadikan data pengujian. Data ini tidak akan digunakan dalam proses pelatihan model, tetapi akan digunakan untuk menguji kinerja model yang telah dilatih.

9. Implementasi Naïve Bayes

Penelitian ini menggunakan algoritma Naive Bayes, yang merupakan salah satu algoritma untuk klasifikasi teks yang didistribusikan secara multinomial. Naive Bayes cocok untuk data teks dengan banyak fitur, seperti kata-kata dalam dokumen, dan menggunakan pendekatan probabilitas untuk klasifikasi. Algoritma ini menghitung probabilitas kemunculan setiap kata dalam setiap kelas untuk memprediksi kelas baru dari suatu teks. Naive

Bayes sering digunakan dalam klasifikasi email, dokumen, dan analisis sentimen teks.

10. Evaluasi

Pada tahap ini, evaluasi dilakukan dengan menggunakan confusion matrix, yang merupakan matriks untuk menghitung jumlah data uji yang dikelompokkan dengan benar dan salah, guna menilai kinerja model klasifikasi. Penggunaan matriks ini membantu dalam menilai kualitas performa model klasifikasi tersebut. Evaluasi ini melibatkan perhitungan accuracy, precision, dan recall. Berikut adalah tabel dari Confusion Matrix:

Tabel 3. 1 Tabel Confusion Matrix

Prediction Class	Actual Class	
	Negative	Positive
Predict Negative	TN	FN
Predict Positive	FP	TP

(Sumber: Penelitian 2024)

Berikut ini persamaan rumus untuk menghitung hasil confusion matrix:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FN + FP)}$$

$$Precision = \frac{TP}{(TP+FP)}$$

$$Recall = \frac{TP}{(TP + FN)}$$

Rumus 3. 2 Persamaan Confusion Matrix

Keterangan rumus:

Accuracy : perbandingan antara jumlah prediksi benar dengan total keseluruhan prediksi.

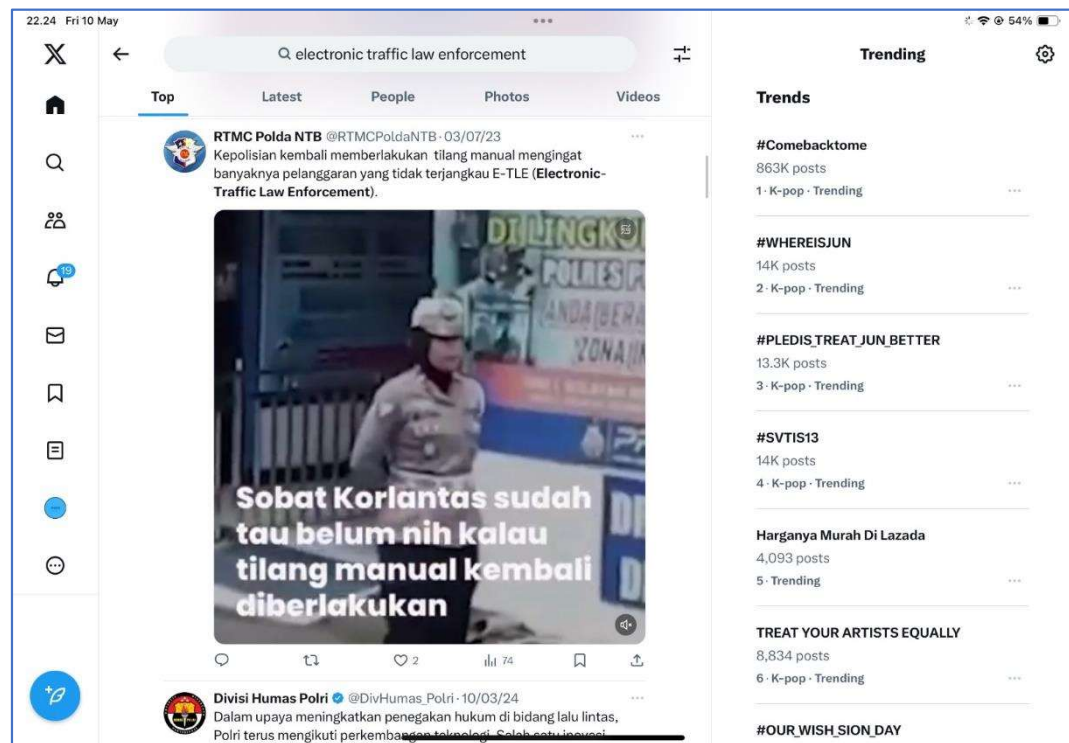
- Precision* : perbandingan antara jumlah prediksi positif yang benar dengan total prediksi positif.
- Recall* : perbandingan antara jumlah prediksi positif yang benar dengan total jumlah data yang benar-benar positif.
- True Positif (TP) : menunjukkan jumlah data yang dengan benar diklasifikasikan sebagai positif.
- True Negatif (TN) : TN menunjukkan jumlah data yang dengan benar diklasifikasikan sebagai negatif
- False Positif (FP) : menunjukkan jumlah data yang salah diklasifikasikan sebagai positif.
- False Negatif (FN) : menunjukkan jumlah data yang salah diklasifikasikan sebagai negatif,

11. Visualisasi

Setelah berhasil menyelesaikan langkah-langkah sebelumnya, tahap akhir melibatkan visualisasi data mengenai sentimen yang diungkapkan terkait penerapan *electronic traffic law enforcement*. Data divisualisasikan melalui *piechart*, *barchart*, dan *wordcloud*. Tujuannya adalah untuk mengevaluasi pendapat publik mengenai polaritasnya, baik itu positif atau negatif. Data teks biasanya mencerminkan subjektivitas yang mengarah pada kesimpulan dari hasil analisis (Vonega et al., 2024).

3.2. Objek Penelitian

Objek dalam penelitian ini adalah opini masyarakat mengenai penerapan peraturan *electronic traffic law enforcement (ETLE)* pada platform media sosial Twitter. Twitter merupakan platform yang efektif untuk mengumpulkan data karena tingkat kebenaran opini yang diunggah (*tweet*) cukup tinggi, sehingga dapat memberikan gambaran tentang pandangan masyarakat terhadap suatu topik. Melalui Twitter, pengguna memiliki kebebasan untuk membagikan berbagai jenis konten seperti opini, sentimen, dan emotikon, yang dapat dijadikan data untuk menganalisis tren atau topik tertentu. Proses analisis data semacam ini dikenal sebagai analisis sentimen atau *opinion mining* (Musfiroh et al., 2021). Memanfaatkan Platform API Twitter, seseorang dapat mengakses beragam data publik Twitter yang tersebar di seluruh dunia. Selain itu, API Twitter juga memberikan dukungan kepada pengguna untuk mengelola privasi mereka dengan memberikan kontrol atas sumber informasi di Twitter, termasuk pesan langsung, yang kemudian dapat diakses oleh pengembang yang telah disetujui (Komara & Hadiapurwa, 2022).



Gambar 3. 2 Tampilan Antarmuka Platform Twitter
(Sumber: Twitter 2024)

Gambar 3.2 merupakan gambar halaman antarmuka platform Twitter. Pada halaman ini ditampilkan data pencarian dengan kata kunci “*electronic traffic law enforcement*”.

3.3. Populasi dan Sampel

Kumpulan dari semua entitas yang menjadi fokus penelitian, termasuk individu dan objek, yang memiliki atribut atau karakteristik tertentu merupakan definisi dari populasi. Dengan demikian, populasi mencakup semua anggota dari kelompok manusia, hewan, peristiwa, atau benda yang berada dalam wilayah tertentu dan menjadi subjek dari kesimpulan yang diambil dari penelitian tersebut (Amin et al., 2023).

Sedangkan sampel merupakan sebagian kecil dari jumlah dan ciri-ciri yang dimiliki oleh populasi. Sampel terdiri dari sejumlah individu yang dipilih dari populasi dan memperlihatkan karakteristik yang mewakili keseluruhan populasi. Sampel yang baik seharusnya mencerminkan populasi dengan baik (Suriani et al., 2023). Maka disimpulkan bahwa, populasi merupakan keseluruhan objek/subjek penelitian, sedangkan sampel merupakan sebagian atau wakil yang memiliki karakteristik representasi dari populasi.

3.3.1 Populasi

Dalam penelitian ini, yang menjadi populasi adalah seluruh data yang diperoleh dari proses *crawling*. Data yang dihasilkan merupakan data acak tanpa memperhitungkan pemilik akun baik secara asli, maupun milik media serta akun yang terindikasi palsu. Total keseluruhan populasi data adalah 1.253 data. Kemudian, data dilakukan tahap *preprocessing* sehingga data akan mudah diolah.

3.3.2 Sampel

Sampel dalam penelitian ini berupa data yang diambil sebagian dari populasi. Lalu, data dibagi ke dalam sentimen, positif dan negatif (proses *labeling data*). Kemudian data dibagi lagi untuk data *training* dan data *testing*. Data yang dijadikan sampel penelitian diambil dengan Teknik non random sampling jenis quota sampling dan purposive sampling. Dengan kedua jenis sampling tersebut ditentukan hanya kluster teks komentar yang dijadikan dataset sampel penelitian dengan total 948 dataset..

3.4. Teknik Pengumpulan Data

Penelitian adalah langkah sistematis yang diambil untuk memahami secara mendalam topik atau fenomena tertentu. Salah satu tahap penting dalam menjalankan penelitian adalah mengumpulkan data yang relevan untuk mendapatkan informasi yang dibutuhkan. Pentingnya teknik pengumpulan data yang sesuai dan instrumen penelitian yang valid sangat besar dalam menghasilkan data yang akurat dan dapat dipercaya (Ardiansyah et al., 2023).

Teknik pengumpulan data merujuk pada prosedur, metode, atau strategi yang digunakan untuk mengumpulkan data yang akan dianalisis. Dengan demikian, teknik ini membutuhkan pendekatan yang terencana dan terstruktur untuk memastikan data yang valid dan ketepatan data yang diperoleh sesuai dengan kondisi sebenarnya (Nashrullah et al., 2023). Penelitian ini dilakukan pengumpulan data dengan cara, observasi dan dokumentasi.

1. **Observasi:** merupakan teknik pengumpulan data yang melibatkan pengamatan langsung terhadap partisipan dan konteks yang terlibat dalam fenomena penelitian. Dalam penelitian ini, observasi dilakukan secara langsung oleh peneliti dengan cara mengunjungi *platform* media sosial Twitter.
2. **Dokumentasi:** melibatkan pengumpulan data dari dokumen, arsip, atau bahan tertulis lainnya yang berkaitan dengan fenomena penelitian. Dalam kasus penelitian ini, dokumen yang berkaitan dikumpulkan di antaranya, jurnal penelitian terdahulu, buku referensi, data-data pendukung penelitian,

dan *database* Twitter yang diperoleh melalui proses *crawling* yang akan dijadikan data utama dalam proses penelitian.

Tahapan atau proses pengumpulan data dari Twitter dimulai dengan menggunakan Python untuk mengambil data dalam proses yang disebut *crawling*. Data yang diperoleh kemudian disimpan dalam format CSV, lalu dilakukan seleksi dan *labeling* agar siap digunakan dalam penelitian.

3.4.1 Crawling Data

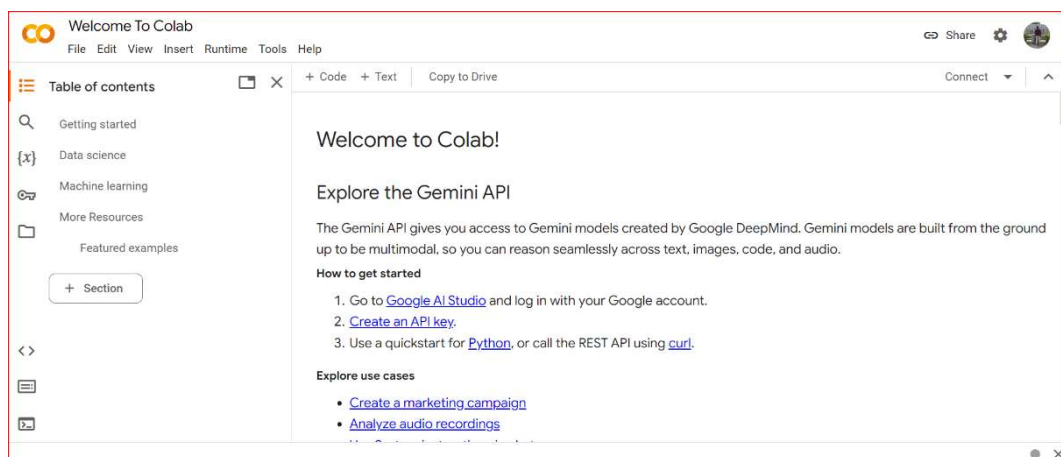
Crawling data merupakan suatu teknik otomatis untuk mengambil data dari sebuah situs web menggunakan program komputer. Pendekatan ini memperbolehkan pengambilan data dari sumber-sumber beragam secara efisien dan terstruktur, sehingga dapat dimanfaatkan untuk berbagai keperluan seperti penelitian, analisis, dan pengembangan aplikasi (Rismawan & Syahidin, 2023). Data dari berbagai media sosial dikumpulkan menjadi satu untuk dievaluasi dan disusun menjadi sebuah penelitian. Proses pengambilan data ini umumnya disebut sebagai *crawling* data (Vindua & Zailani, 2024).

Dalam penelitian ini, proses *crawling* data dari Twitter memanfaatkan *Twitter Auth Token* yang diperoleh dari akun Twitter peneliti. Untuk kode program menggunakan bahasa pemrograman Python yang dituliskan dengan bantuan *platform* Google Colab. Google Colaboratory atau Colab merupakan, sebuah *executable document* yang dapat digunakan untuk menyimpan, menulis, serta membagikan program yang telah ditulis melalui Google Drive. *Software* ini pada dasarnya adalah layanan *Jupyter Notebook* yang di-*hosting* dan dapat digunakan

tanpa penyiapan, serta menyediakan akses tanpa biaya ke *resource* komputasi, termasuk GPU dan TPU. Colab sangat cocok untuk *machine learning*, ilmu data, dan pendidikan.

Berikut ini tahapan crawling data yang dilakukan oleh peneliti melalui platform Google Colab. Kode pemrograman menggunakan Python yang diperoleh melalui internet.

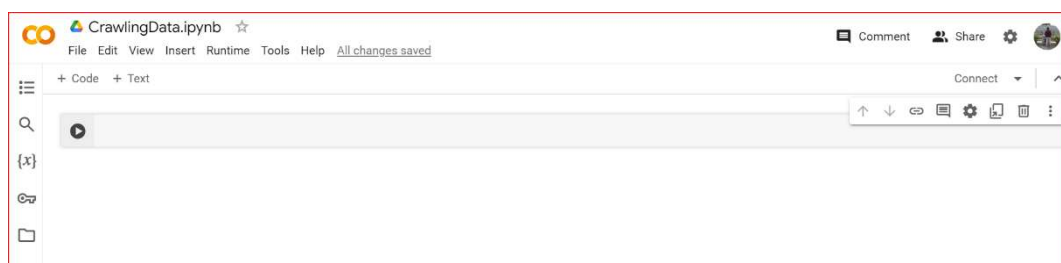
Langkah 1. Membuka halaman web Google Colab dan *login* akun gmail.



Gambar 3. 3 Halaman Depan Google Colab
(Sumber: Google Colab 2024)

Gambar 3.3 merupakan tampilan halaman utama Software Google Colab yang diakses melalui peramban Google Chrome.

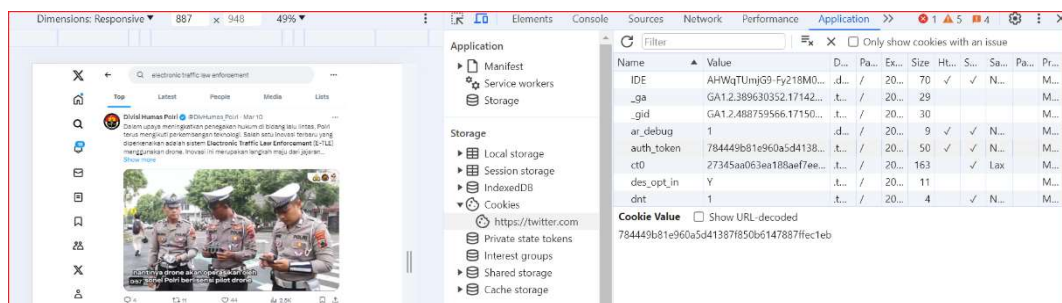
Langkah 2. Membuat *file* dokumen kerja baru.



Gambar 3. 4 Lembar Dokumen Baru
(Sumber: Google Colab 2024)

Gambar 4.4 merupakan tampilan lembar kerja baru, untuk menuliskan kode program Python.

Langkah 3. Mengambil kode autentikasi Twitter, dengan cara *inspect* web Twitter.



Gambar 3. 5 Tampilan Inspect Akun Twitter
(Sumber: Twitter 2024)

Gambar 3.5 merupakan proses pengambilan kode autentikasi akun Twitter melalui web. Proses ini dilakukan dengan cara membuka halaman inspect pada website.

Langkah 4. Menuliskan kode Python berikut dengan kode autentikasi yang telah diambil dari Twitter.

```
#@title Twitter Auth Token
twitter_auth_token = '784449b81e960a5d41387f850b6147887ffec1eb'
```



Gambar 3. 6 Implementasi Auth Token Twitter
(Sumber: Google Colab 2024)

Gambar 3.6 merupakan proses penulisan kode autentikasi akun twitter ke dalam halaman kerja Google Colab.

Langkah 5. Melakukan instalasi *package* Python dengan menuliskan kode berikut.

```
#Import required Python package
```

```

!pip install pandas

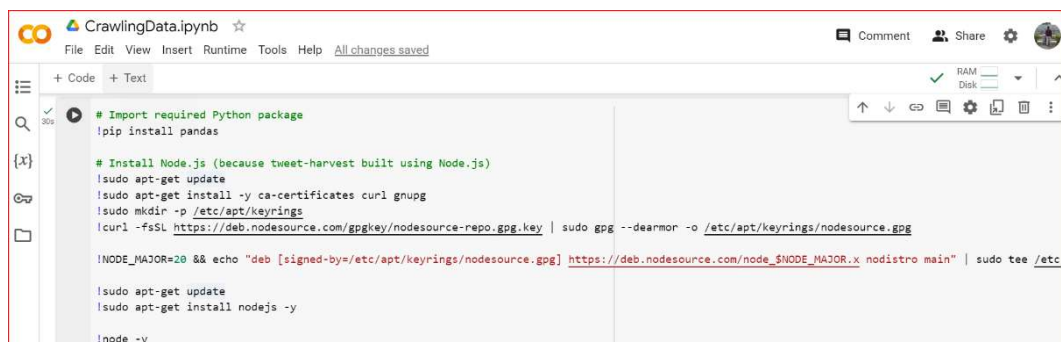
# Install Node.js (because tweet-harvest built using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg

!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_${NODE_MAJOR}.x nodistro main" | sudo tee /etc/apt/sources.list.d/nodesource.list

!sudo apt-get update
!sudo apt-get install nodejs -y

!node -v

```



Gambar 3. 7 Instalasi Package Python
(Sumber: Google Colab 2024)

Gambar 3.7 merupakan proses instalasi file package library Python pada lembar kerja Google Colab.

Langkah 6. Menuliskan *filename* penyimpanan *dataset*, kata kunci pencarian Tweet yang dibutuhkan dan jumlah limit data yang dibutuhkan. Berikut kode programnya.

```

filename = 'ETLE.csv'
search_keyword = 'Electronic Traffic law Enforcement lang:id'

```

```

limit = 500

!npx -y tweet-harvest@2.6.0 -o "{filename}" -s
"{search_keyword}" --tab "LATEST" -l {limit} --token
{twitter_auth_token}

```



Gambar 3. 8 Penulisan Kata Kunci Tweet
(Sumber: Google Colab 2024)

Gambar 3.8 merupakan proses penulisan kata kunci pencarian data tweet. Kata kunci yang digunakan adalah “electronic traffic law enforcement, ETLE, e-tilang dan Tilang elektronik”.

Langkah 7. untuk menampilkan hasil *crawling* data, dibutuhkan library pandas.

Berikut kode program yang dibutuhkan.

```

#import library
import pandas as pd

#Specify the path to your CSV file
file_path = f"tweets-data/{filename}"

#Read the CSV file into a pandas DataFrame
df = pd.read_csv(file_path, delimiter=",")

#Display the DataFrame
display(df)

```



```

CrawlingData.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
# import library
import pandas as pd

# Specify the path to your CSV file
file_path = f"tweets-data/{filename}"

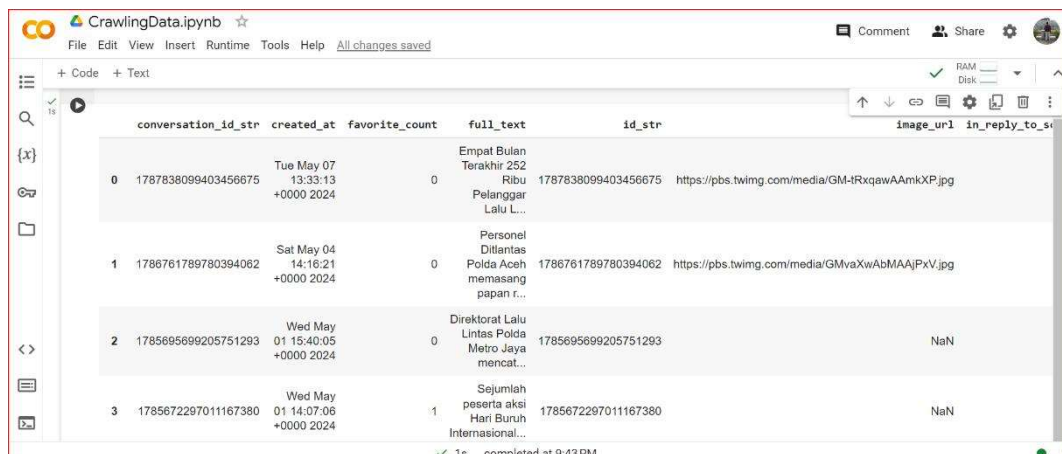
# Read the CSV file into a pandas DataFrame
df = pd.read_csv(file_path, delimiter=",")

# Display the DataFrame
display(df)

```

Gambar 3. 9 *Import Library* Untuk Menampilkan Data
(Sumber: Google Colab 2024)

Gambar 3.9 merupakan proses untuk *mengimport library pandas* digunakan untuk menampilkan data. Tampilan dari data tersebut bisa dilihat pada gambar 3.10.



	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_status_id_str
0	1787838099403456675	Tue May 07 13:33:13 +0000 2024	0	Empat Bulan Terakhir 252 Ribu Pelanggaran Lalu L...	1787838099403456675	https://pbs.twimg.com/media/GM-HRxqawAAAmkXP.jpg	
1	1786761789780394062	Sat May 04 14:16:21 +0000 2024	0	Personel Dillantas Polda Aceh memasang papan r...	1786761789780394062	https://pbs.twimg.com/media/GMvaXwAbMAAJPxV.jpg	
2	1785695699205751293	Wed May 01 15:40:05 +0000 2024	0	Direktorat Lalu Lintas Polda Metro Jaya mencatat...	1785695699205751293		NaN
3	1785672297011167380	Wed May 01 14:07:06 +0000 2024	1	Sejumlah peserta aksi Hari Buruh Internasional...	1785672297011167380		NaN

Gambar 3. 10 Hasil *Crawling* Data
(Sumber: Google Colab 2024)

Gambar 3.10 merupakan tampilan dataset dari Twitter, data tersebut secara otomatis tersimpan ke dalam file CSV yang dijadikan sebagai dataset.

3.4.2. Dataset

Dataset yang didapatkan merupakan opini yang dituliskan dalam bahasa Indonesia. Pencarian *tweet* menggunakan kata kunci “*electronic traffic law enforcement*”. Karena data yang didapatkan kurang, maka kata kunci “ETLE”, “E-

Tilang” dan “Tilang Elektronik” ditambahkan guna menambah jumlah data yang dibutuhkan. Berikut ini adalah tabel hasil dari *crawling* data yang didapatkan:

Tabel 3. 2 Hasil Dataset Crawling Data

No	Kata Kunci	Jumlah Data (Raw)
1	“Electronic Traffic Law enforcement”	142 Data
2	“ETLE”	95 Data
3	“E-Tilang”	513 Data
4	“Tilang Elektronik”	503 Data
5	Total	1253 Data

(Sumber: Penelitian 2024)

conversation_id_str	created_at	favorite_count	full_text	id_str	image_url_in_reply_to_tweet	location	quote_count	reply_count	retweet_count	tweet_url	user_id	username	
1	1.78E+18 Thu Apr 21	0	Halo Sana	1.78E+18	https://pbs.twimg.com	Kalimanta	0	0	0	https://tw.144E+18	pemkabkukar		
2	1.78E+18 Wed Apr 2	6	Siapa di si	1.78E+18	https://pbs.twimg.com	DKI Jakart	0	2	4	https://tw.76647722	TMCPoldaMetro		
3	1.78E+18 Thu Apr 1	0	Sebanyak	1.78E+18	https://pbs.twimg.com		0	1	0	https://tw.57261519	Metro_TV		
4	1.78E+18 Thu Apr 1	1	Siapa di si	1.78E+18	https://pbs.twimg.com	DKI Jakart	0	2	2	https://tw.76647722	TMCPoldaMetro		
5	1.77E+18 Mon Apr 0	2	Siapa di si	1.77E+18	https://pbs.twimg.com	DKI Jakart	0	0	1	https://tw.76647722	TMCPoldaMetro		
6	1.77E+18 Sun Mar 3	3	Siapa di si	1.77E+18	https://pbs.twimg.com	DKI Jakart	0	0	3	https://tw.76647722	TMCPoldaMetro		
7	1.77E+18 Sat Mar 3	1	Siapa di si	1.77E+18	https://pbs.twimg.com	DKI Jakart	1	0	0	https://tw.76647722	TMCPoldaMetro		
8	1.77E+18 Fri Mar 29	0	Satlantas	1.77E+18		Malang, Ir	0	0	0	https://tw.2.81E+08	kalimaya_malang		
9	1.77E+18 Thu Mar 2	3	Siapa di si	1.77E+18	https://pbs.twimg.com	DKI Jakart	0	1	1	https://tw.76647722	TMCPoldaMetro		
10	1.77E+18 Thu Mar 2	7	Siapa di si	1.77E+18	https://pbs.twimg.com	DKI Jakart	0	2	2	https://tw.76647722	TMCPoldaMetro		
11	1.77E+18 Tue Mar 2	0	Petugas ti	1.77E+18		To RP	0	1	0	https://tw.2.23E+09	bbxayi		
12	1.77E+18 Tue Mar 2	1	16. Petuga	1.77E+18	https://pt	Sudastika	in	Under the	0	1	1	https://tw.3.41E+08	Sudastika
13	1.77E+18 Fri Mar 22	4	Siapa di si	1.77E+18	https://pbs.twimg.com	DKI Jakart	0	1	0	https://tw.76647722	TMCPoldaMetro		
14	1.77E+18 Thu Mar 2	0	Direktur P	1.77E+18	https://pbs.twimg.com		0	0	0	https://tw.1.45E+18	Suroboy96849901		
15	1.77E+18 Thu Mar 2	0	Direktur P	1.77E+18	https://pbs.twimg.com		0	0	0	https://tw.1.45E+18	Suroboy96849901		
16	1.77E+18 Thu Mar 2	0	Direktur P	1.77E+18	https://pbs.twimg.com		0	0	0	https://tw.1.45E+18	Suroboy96849901		
17	1.77E+18 Thu Mar 2	0	Direktur P	1.77E+18	https://pbs.twimg.com		0	0	0	https://tw.1.45E+18	Suroboy96849901		
18	1.77E+18 Thu Mar 2	0	Direktur P	1.77E+18	https://pbs.twimg.com		0	0	0	https://tw.1.45E+18	Suroboy96849901		
19	1.77E+18 Thu Mar 2	0	Direktur P	1.77E+18	https://pbs.twimg.com		0	0	0	https://tw.1.45E+18	Suroboy96849901		
20	1.77E+18 Thu Mar 2	0	Direktur P	1.77E+18	https://pbs.twimg.com		0	0	0	https://tw.1.45E+18	Suroboy96849901		
21	1.77E+18 Thu Mar 2	0	Direktur P	1.77E+18	https://pbs.twimg.com		0	0	0	https://tw.1.45E+18	Suroboy96849901		

Gambar 3. 11 Data Yang Disimpan Dalam Format CSV

(Sumber: Penelitian 2024)

Dataset yang telah didapatkan, dilakukan seleksi data dengan hanya mengambil *full text tweet*. Terdapat 1.253 total *dataset* yang didapatkan dari keseluruhan jumlah data. *Dataset* tersebut masih bisa berubah setelah melalui preprocessing data. Contoh *dataset* ditunjukkan pada Tabel 3.2 berikut.

Tabel 3. 3 Dataset Teks Tweet

No	Teks
1	Terima kasih @TMCPoldaMetro yg sdh memasang CCTV E-TILANG di kawasan sekitar Pasar Pondok Labu yang sering pengguna motor lawan arah mohon di Jl Bungur Tanah Kusir & dibawah flyover Ciputat dipasang CCTV E-tilang juga biar pada tertib terima kasih. @RadioElshinta @detikcom https://t.co/TYgetvtLeA

2	Yang penting adalah tolong tiap naik mobil duduk sebelah driver inisiatif pasang sabuk pengaman. Lu yg gamau pasang kita yg kena e-tilang anjrit. Mau marah.
3	Tapi emang kayanya tuh pakpol nyari duit di hari minggu soalnya lgsg patok harga cepe katanya biar plat motor kamu di blokir dari kamera cctv (jadi ga kena e-tilang). Au dah tuh bener apa kagak yang penting gue ga sidang aja. Maless urusannya soalnya ketilang di jakpus
4	Ke rakyat metode pengawasannya canggih betul pake cctv hingga e-tilang Giliran ke instansi internal modal bacot dan rakyat wajib percaya
5	@kegblgnunfaedh @medyrenaldy_ Udah dari dulu kyk gini percuma gak ada perubahan
6	@convomf suka memfoto diam2 Mo nikah sma kamera e-tilang??
7	Di persimpangan Jl RS Fatmawati & Jl Kartini banyak pengguna kendaraan bermotor tidak tertib lalu lintas mohon dipasang kamera E-tilang agar bisa tertib & juga petugas Polisi Lalulintas utk berjaga mentertibkan Tx. @RadioElshinta @TMCPoldaMetro @Poldametrojaya @DivHumas_Polri https://t.co/IhGhyNfWp5
8	Nama di STNK ini kan juga berlaku buat E-tilang. Misal nama kita dipake buat kendaraan yang dipakai kejahatan emang rela? Belum lagi kena pajak progresif padahal di rumah cuma ada satu kendaraan karena namamu dipake orang
...
1253	@RadioElshinta @TMCPoldaMetro Maaf ya mnrt sya separator sepanjang cawang - polda metro ga perlu krn sdh ada e-tilang. Tp klo pun hrs ada setidaknya di review.

(Sumber: Data Penelitian 2024)

3.5. Model Penelitian

Penggunaan *electronic traffic law enforcement* (ETLE) yang diusulkan oleh Korlantas Polri, telah menimbulkan berbagai reaksi dari masyarakat. Beragam reaksi tersebut diutarakan melalui media sosial, salah satunya adalah *platform* Twitter. Penelitian ini akan memanfaatkan data *tweet* pengguna Twitter yang didapatkan dengan cara *crawling data*. Data *tweet* yang didapatkan akan diubah ke dalam format CSV untuk dijadikan *dataset* penelitian. *Dataset* akan melalui tahap pelabelan data sentimen dan *preprocessing* data agar data siap untuk diolah menggunakan model algoritma Naive Bayes. Tujuan penelitian ini adalah untuk

menganalisis sentimen masyarakat terhadap pro dan kontra terkait penerapan ETLE. Hasil penelitian akan divisualisasikan ke dalam bentuk diagram dan *wordcloud*.