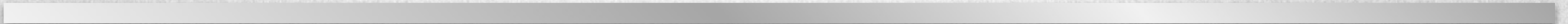


[illegible]



Data Acquisition

- **Data Acquisition** merupakan nama lain dari pengumpulan data.



Sumber Data

- Data Primer
 - ❖ Survey, Sensus
 - ❖ Eksperimen/Percobaan
 - ❖ Scraping, Crawling
- Data Skunder : data yang bersumber dari pihak ketiga

Scraping dan Crawling

- **Crawling** merupakan teknik mengumpulkan data pada sebuah website dengan memasukkan **Uniform Resource Locator (URL)**
- **Crawling** biasanya menggunakan **Application Programming Interface (API)** sebagai jalur komunikasi dalam mendapatkan data.
- **Scraping** merupakan teknik mengumpulkan data pada sebuah website melalui proses ekstraksi informasi menggunakan **Hypertext Transfer Protocol (HTTP)**.
- **Scraping** biasanya perlu mengetahui **element HTML** ataupun **XML** pada sebuah website

Legalitas Scraping

- Scrapping secara umum (minimal) **tidak etis** (bisa jadi ilegal). Mengapa? karena scraping mirip **DDOS attack** yang akan memberatkan server atau bahkan membuat server berhenti berfungsi normal.
- Program scraping juga memungkinkan akan mengakses data yang **tidak dimaksudkan untuk konsumsi publik**.
- Satu-satunya saat dimana scraping **boleh dilakukan** adalah saat sang pelaku/programmer menghormati “**robots.txt**” yang telah diberikan oleh web administrator.
- Menggunakan **API (Application Program Interface)** yang diberikan oleh provider (website/medsos) lalu melakukan crawling.
- Yakinkan untuk membaca **ToS (Terms of Service)** dengan baik.

Twitter Data Acquisition

- **Crawling** menggunakan package **tweepy** di **python**
- Tweepy merupakan package python untuk mengakses **API Twitter resmi**
- **Tweepy** perlu **mendaftar** sebagai **developer** dan dan mendapatkan **kunci** dari **Twitter**
- Scrapping menggunakan package **twint** di **python**
- **Twint** tidak menggunakan **API Twitter Resmi**

See More:

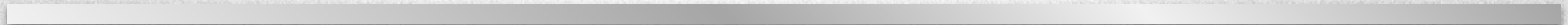
- Tweepy documentation: <https://docs.tweepy.org/en/latest/>
- Twint documentation: <https://github.com/twintproject/twint/wiki>

tweepy vs twint

- **Performa** : **twint** jauh lebih lambat untuk mendapatkan data dibandingkan **tweepy**
- **Limitasi** : **tweepy** bergantung pada limitasi **Twitter API**, seperti jumlah tweet yang bisa dikumpulkan terbatas.

tweepy

- Untuk bisa menggunakan package tweepy kita harus mendaftar terlebih dahulu di Twitter Developer
- Dari website tersebut akan didapatkan semacam **key** yang harus dimasukan ke program

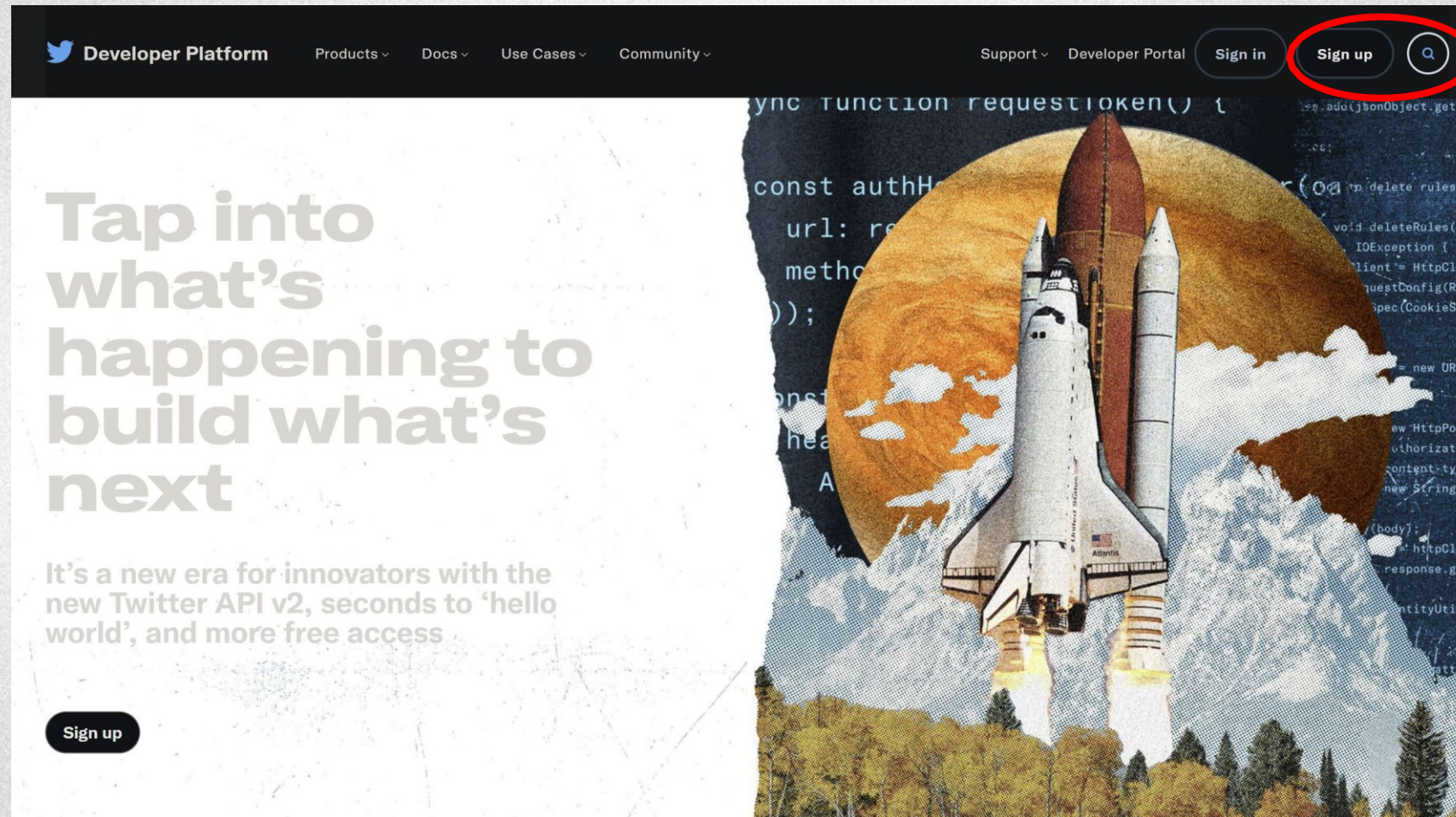


Tutorial Daftar **Twitter** developer

Langkah-langkah untuk mendapatkan key dan token dari twitter:

1. Login ke twitter (buat bila belum ada dengan menyertakan nomor hp)
2. Registrasi sebagai developer: <https://developer.twitter.com/en>
3. Klik Sign Up

Tutorial Daftar **Twitter** developer



Tutorial Daftar **Twitter** developer

4. Isi form sesuai dengan kebutuhan
5. Klik **Let's do this**

Twitter Account ⓘ

@ gad_auth

[Switch @username](#) [Create new @username](#)

Email ⓘ

zo****@gm****.com

[Change email](#)

What country are you based in?

Indonesia

What's your use case?

We need this information for data protection. [Learn more](#)

Academic researcher

Will you make Twitter content or derived information available to a government entity or a government affiliated entity? ⓘ

No

Get involved and updated

☐ Sign me up for the latest Twitter Dev news, tips and updates

[Need help?](#)

Let's do this

Tutorial Daftar **Twitter** developer

6. Centang Accept **Terms & Conditions**

7. Klik **Submit**

Docs ▾ Communi

Developer agreement & policy

Developer Agreement

Effective: January 19, 2023

This Twitter Developer Agreement (“**Agreement**”) is made between you (either an individual or an entity, referred to herein as “**you**”) and Twitter (as defined below) and governs your access to and use of the Licensed Material (as defined below). Your use of Twitter’s websites, SMS, APIs, email notifications, applications, buttons, embeds, ads, and our other covered services is governed by our general Terms of Service and Privacy Policy.

PLEASE READ THE TERMS AND CONDITIONS OF THIS AGREEMENT CAREFULLY, INCLUDING ANY LINKED TERMS REFERENCED BELOW, WHICH ARE PART OF THIS LICENSE AGREEMENT. BY USING THE LICENSED MATERIAL, YOU ARE AGREEING THAT YOU HAVE READ,

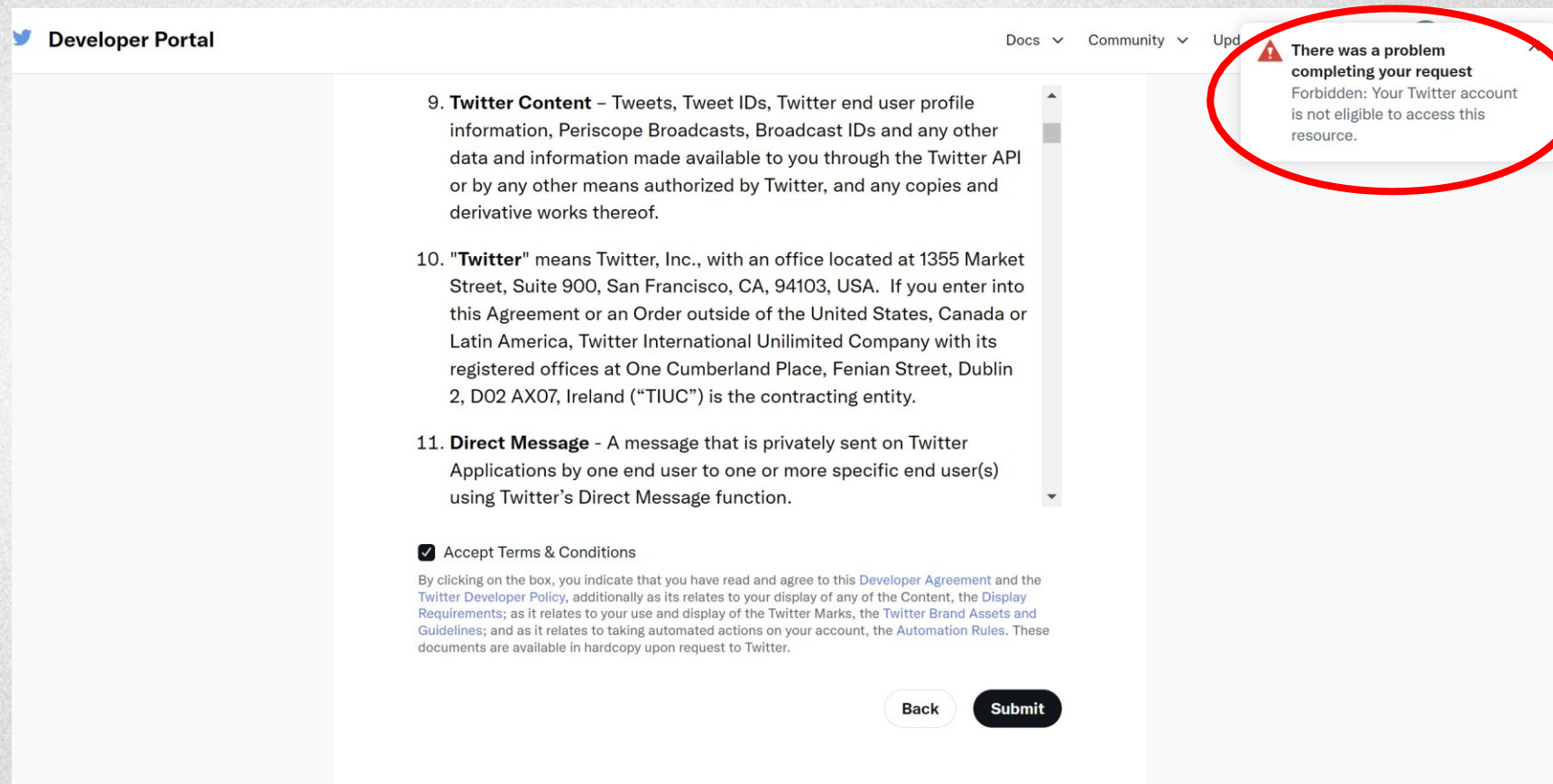
☒ Accept Terms & Conditions

By clicking on the box, you indicate that you have read and agree to this [Developer Agreement](#) and the [Twitter Developer Policy](#), additionally as it relates to your display of any of the Content, the [Display Requirements](#); as it relates to your use and display of the Twitter Marks, the [Twitter Brand Assets and Guidelines](#); and as it relates to taking automated actions on your account, the [Automation Rules](#). These documents are available in hardcopy upon request to Twitter.

BackSubmit

Tutorial Daftar **Twitter** developer

Jika belum menyertakan nomor telpon di akun twiiter maka akan muncul error



The screenshot shows the Twitter Developer Portal registration form. The form includes sections for terms and conditions, a checkbox for "Accept Terms & Conditions", and "Back" and "Submit" buttons. An error message is displayed in the top right corner, circled in red, indicating a problem with the user's Twitter account.

Developer Portal

9. **Twitter Content** – Tweets, Tweet IDs, Twitter end user profile information, Periscope Broadcasts, Broadcast IDs and any other data and information made available to you through the Twitter API or by any other means authorized by Twitter, and any copies and derivative works thereof.

10. **"Twitter"** means Twitter, Inc., with an office located at 1355 Market Street, Suite 900, San Francisco, CA, 94103, USA. If you enter into this Agreement or an Order outside of the United States, Canada or Latin America, Twitter International Unilimited Company with its registered offices at One Cumberland Place, Fenian Street, Dublin 2, D02 AX07, Ireland ("TIUC") is the contracting entity.

11. **Direct Message** - A message that is privately sent on Twitter Applications by one end user to one or more specific end user(s) using Twitter's Direct Message function.

☒ Accept Terms & Conditions

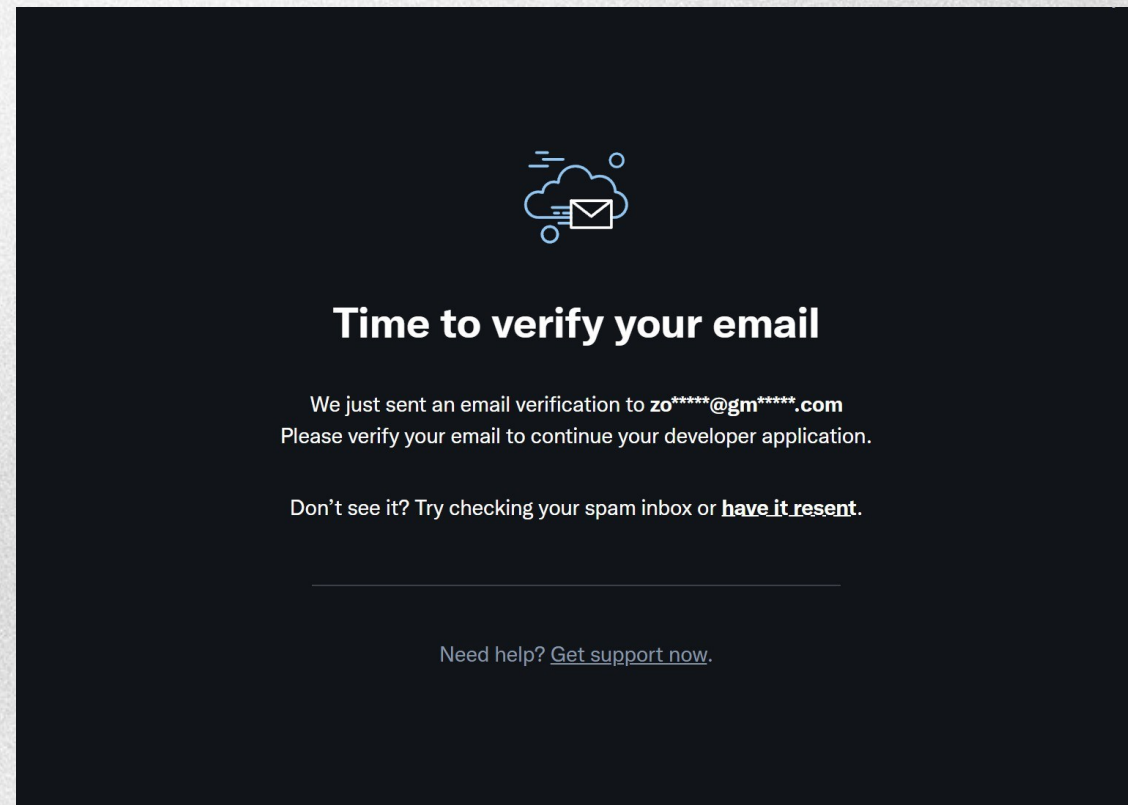
By clicking on the box, you indicate that you have read and agree to this [Developer Agreement](#) and the [Twitter Developer Policy](#), additionally as it relates to your display of any of the Content, the [Display Requirements](#); as it relates to your use and display of the Twitter Marks, the [Twitter Brand Assets and Guidelines](#); and as it relates to taking automated actions on your account, the [Automation Rules](#). These documents are available in hardcopy upon request to Twitter.

[Back](#) [Submit](#)

There was a problem completing your request
Forbidden: Your Twitter account is not eligible to access this resource.

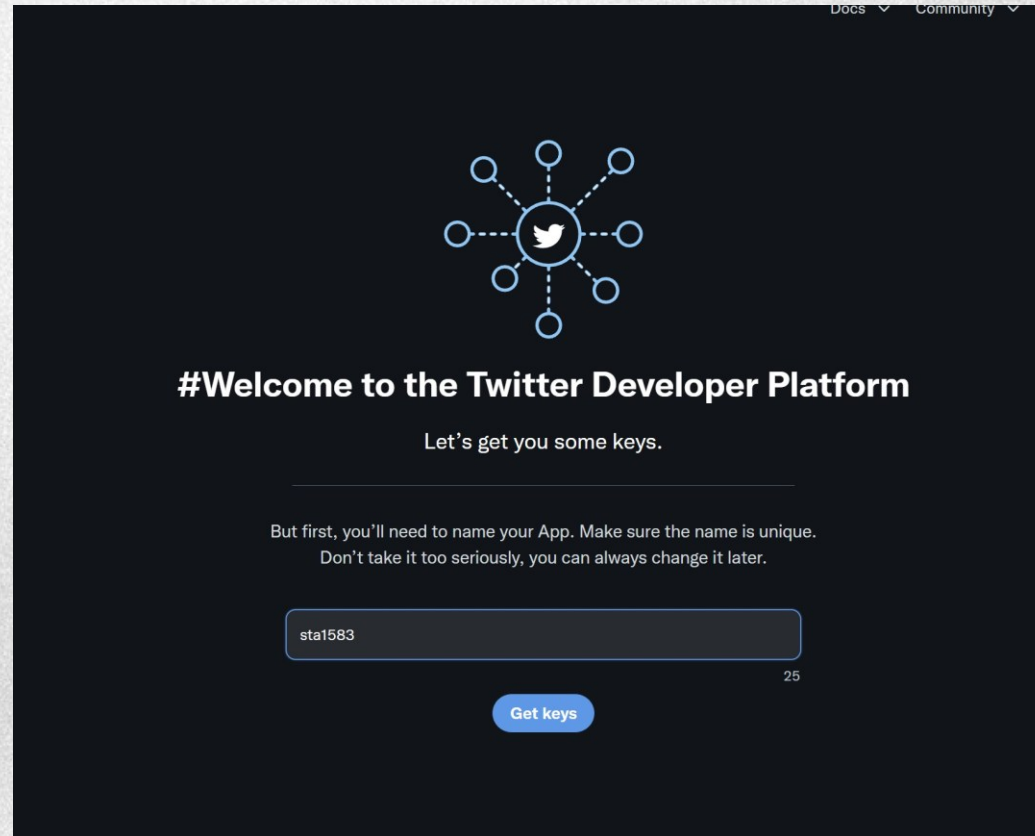
Tutorial Daftar **Twitter** developer

8. Periksa Email anda dan kemudian konfirmasi melalui email anda



Tutorial Daftar **Twitter** developer

9. Tuliskan nama aplikasi anda untuk mendapatkan key:



#Welcome to the Twitter Developer Platform

Let's get you some keys.

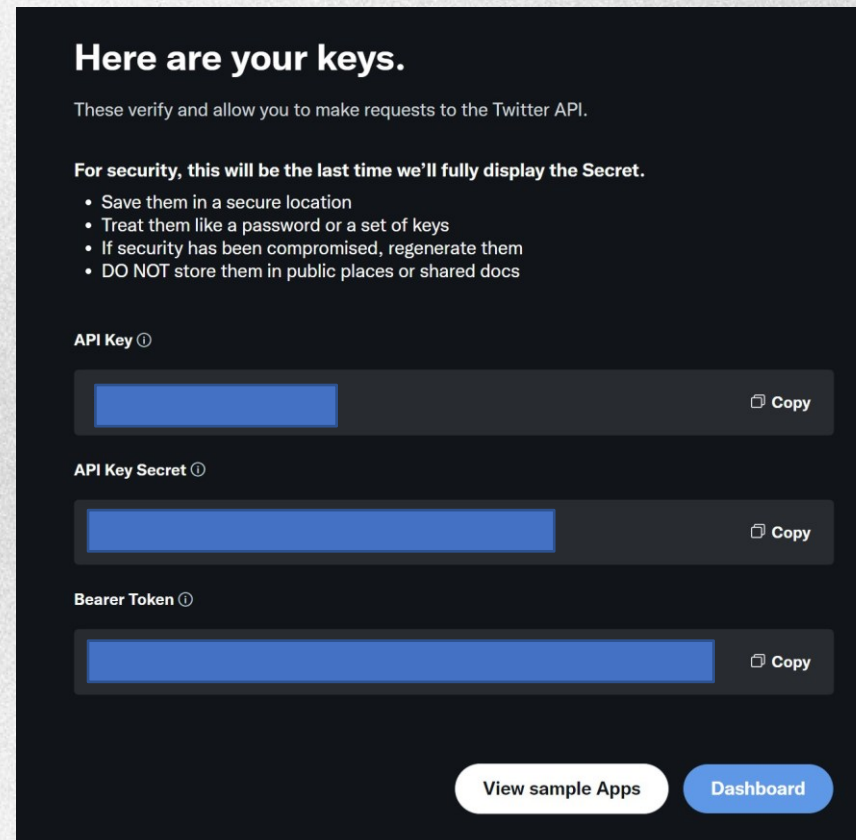
But first, you'll need to name your App. Make sure the name is unique.
Don't take it too seriously, you can always change it later.

sta1583

Get keys

Tutorial Daftar **Twitter** developer

10. Simpan informasi ini (api key, api key secret, bearer token) di tempat yang aman (notepad/screenshoot/dll)



Here are your keys.

These verify and allow you to make requests to the Twitter API.

For security, this will be the last time we'll fully display the Secret.

- Save them in a secure location
- Treat them like a password or a set of keys
- If security has been compromised, regenerate them
- DO NOT store them in public places or shared docs

API Key ⓘ

Copy

API Key Secret ⓘ

Copy

Bearer Token ⓘ

Copy

[View sample Apps](#) [Dashboard](#)

Tutorial Daftar **Twitter** developer

11. Klik dashboard
12. Klik Yes, I saved them
13. Selesai

Did you save your Keys and Tokens?

- Save them in a secure location
- Treat them like a password or a set of keys
- If security has been compromised, regenerate them
- DO NOT store them in public places or shared docs

API Key ⓘ

[Copy](#)

API Key Secret ⓘ

[Copy](#)

Bearer Token ⓘ

[Copy](#)

[Cancel](#) [Yes, I saved them](#)

Search queries untuk **Twitter**

- Search queries digunakan untuk mendapatkan tweet yang kita inginkan
- Search queries terdiri dari keyword dan konjungsi-operator seperti **is:retweet**

Search queries untuk **Twitter**

- Contoh Search queries: **from:pemersatubangsa** berarti kita akan mendapatkan semua tweet dari akun **pemersatubangsa**
- Contoh Search queries: **corona coronavirus covid-19** berarti kita akan mendapatkan tweet yang memuat kata **corona, coronavirus dan covid-19**.
“Spasi” diantara kata berarti operator logika AND.
- Contoh Search queries: **from:pemersatubangsa -is:retweet** berarti kita akan mendapatkan semua tweet dari akun **pemersatubangsa** tanpa **retweet**. Operator **is:retweet** berarti me-filter tweet yang berupa retweet dan operator “-” berarti menegasikan kondisi.

Search queries untuk **Twitter**

- Contoh Search queries: **covid-19 has:geo** berarti kita akan mendapatkan tweet yang memuat kata **covid-19** dan terdapat data **geo** didalamnya
- Contoh Search queries: **from:pemersatubangsa has:images** berarti kita akan mendapatkan semua tweet dari **pemersatubangsa** yang terdapat gambar didalamnya.
- Contoh Search queries: **covid-19 lang:en** berarti kita akan mendapatkan tweet yang memuat kata **covid-19** dan berbahasa inggris.

Search queries untuk **Twitter**

Operator	Description
\$	Matches any Tweet that contains the specified 'cashtag'.
place	Matches Tweets tagged with the specified location or Twitter place ID
place_country	Matches Tweets where the country code associated with a tagged place/location matches the given ISO alpha-2 character code
point_radius	Matches against the place.geo.coordinates object of the Tweet when present, and in Twitter, against a place geo polygon, where the Place polygon is fully contained within the defined region
bounding_box	Matches against the place.geo.coordinates object of the Tweet when present, and in Twitter, against a place geo polygon, where the place polygon is fully contained within the defined region
-is:nullcast	removes Tweets created for promotion only on ads.twitter.com that have a "source":"Twitter for Advertisers (legacy)" or "source":"Twitter for Advertisers"
has:cashtags	Matches Tweets that contain a cashtag symbol (with a leading '\$' character. For example, \$tag)
has:geo	Matches Tweets that have Tweet-specific geolocation data provided by the Twitter user
bio	Matches a keyword within the Tweet publisher's user bio name
bio_name	Matches a keyword within the Tweet publisher's user bio name
bio_location	Matches Tweets that are published by users whose location contains the specified keyword or phrase

Referensi

- [Teknik Pengumpulan Data Sekunder dengan Web Crawling \(dqlab.id\)](#)
- [Perbedaan Antara Crawling dan Scraping | by Dede Brahma | Medium](#)
- [taudata Analytics: Apakah scraping website/media sosial diperbolehkan?](#)
- [What Python package is best for getting data from Twitter? Comparing Tweepy and Twint. | by JP Hwang | Towards Data Science](#)
- [<https://github.com/twitterdev/getting-started-with-the-twitter-api-v2-for-academic-research/blob/main/modules/5-how-to-write-search-queries.md>](#)
- [How to Access Data from the Twitter API using Tweepy \(Python\) | by Michael Galarnyk | Towards Data Science](#)