Analisa *Tweet* Dengan *Keyword Tweet* Saham

**Muhammad Asyroful Nur Maulana Yusuf 1**

1 Teknik Informatika, Institut Teknologi Sumatera  
[muhammad.11914002](mailto:muhammad.11914002)6@student.itera.ac.id

|  |  |  |
| --- | --- | --- |
| **Article Info** |  | **ABSTRACT** |
| ***Article history:***  Received 28 Maret 2022  Revised …  Accepted … |  | Microblogging social media services that are very popular by users on the internet such as twitter. Twitter is a place for users to tweet on short sentences with a wide variety of keywords. Stocks are becoming a popular conversation on twitter because of their association with the capital market. Data retrieval on twitter can be done by automated crawling methods on the internet. Raw data needs to use the preprocessing stage to filter out unnecessary data. The information studied from preprocessing data can be done by visualization in an informative and interesting manner. The information studied from preprocessing data can be done by visualization in an informative and interesting manner. The information studied from preprocessing data can be done by visualization in an informative and interesting manner.The information studied from preprocessing data can be done by visualization in an informative and interesting manner. The information studied from preprocessing data can be done by visualization in an informative and interesting manner. |
| ***Keyword:***  *Crawling, Preprocessing, Saham, Tweet, Visualisasi* |
|  |

**I. Pendahuluan**

Pada saat ini situs *microblogging* telah menjadi alat komunikasi yang sangata popular di internet. *Microblogging* merupakan suatu layanan media sosial yang memungkinkan pengguna mem-publish pesan pendek berupa opini, komentar, berita dalam karakter yang terbatas (kurang dari 200 karakter) salah satu contoh layanan *microblogging* adalah *twitter.*[1] *Twitter* memungkinkan pengguna untuk berbagi pesan menggunakan teks pendek disebut *Tweet*. *Tweet* seringkali digunakan untuk mengungkapkan suatu hal dalam bentuk apapun ataupun kata kunci seperti Saham.

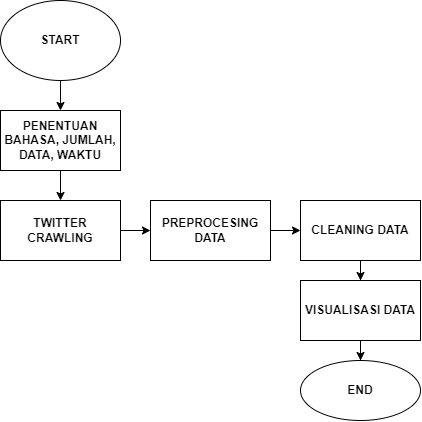
Saham adalah merupakan salah satu produk pasar modal berbentuk instrumen pasar yang paling popular saat ini. Saham disebut efek atau stock menjadi surat berharga yang menyatakan bukti kepemilikan sebuah perusahaan atau badan usaha. Saham dapat diartikan sebagai tanda penyertaan modal perseorangan atau badan usaha sebagai pihak tertentu dalam perusahaan atau perseroan terbatas.[2] Berdasarkan UU Pasar Modal Nomor. 8 / 1995 pasal 1 ayat 5 adalah surat berharga yang menjadi surat pengakuan utang, surat berharga komersial, saham, obligasi, unit penyertaan kontrak investasi kolektif, tanda buku utang, kontrak berjangka, dan bentuk derivatif.[3]

Analisa *tweet* dapat dilakukan dengan berbagai macam, salah satunya adalah menganalisa *tweet* secara menyeluruh berdasarkan waktu t*weet,* kalimat *tweet,* jumlah *favorite tweet,* jumlah *retweets,* dan *device* yang digunakan pengguna dalam melakukan *Tweet*. Dimana hasil dari Analisa tersebut nantinya akan divisualisasikan sebagai informasi yang dikaji secara informatif dan interaktif.

Menurut Mckinlay Shneiderman, definisi visualisasi adalah penggambaran data secara visual yang interaktif dengan menggunakan teknologi sebagai alat pendukung untuk memperkuat hasil pengamatan ataupun Analisa.[4] Menurut Mc Cormick, visualisasi informasi memiliki empat karakteristik diantaranya menggunakan pola, perbandingan gambar, gambar animasi, dan warna. Penggunaan pola berfungsi untuk menyimpulkan informasi secara cepat berdasarkan pola-pola yang membedakan pola satu dengan pola lainnya.[5] Perbandingan gambar dapat berupa bentuk yang merupakan pembeda antara visual yang satu dengan yang lain. Animasi dapat menggambarkan ataupun membedakan visualisasi berdasarkan perjalannya waktu. Deskripsi warna dengan warna yang berbeda dapat mempengaruhi perbedaan informasi yang dihasilkan.[6]

Analisa *tweet* yang sudah divisualisasikan dapat memberikan manfaat bagi pembaca yaitu dapat mengetehaui jumlah *tweet, retweets, favorite, user status* berdasarkan jangka waktu tertentu, kata yang sering muncul pada *tweet* dengan kata kunci Saham, jumlah t*weet, retweets, favorite, user status* berdasarkan *device* yang digunakan pengguna ketika *tweet* dengan kata kunci saham.

**II. METODE/ALUR PENGAMBILAN DATA**



Gambar 1. Alur Pengambilan dan Pemrosesan Data

Pada gambar di atas merupakan alur *Crawling Twitter* sampai Visualisasi Data, diamana prosesnya diawali dengan penentuan Bahasa *tweet,* jumlah *tweet,* dan waktu *tweet* pada *Twitter* untuk proses analisis. Proses pengambilan data digunakan dengan menggunakan metode *crawling*, dimana *crawling* merupakan metode yang digunakan untuk mengambil data secara otomatis dari server twitter berupa user dan tweet beserta atibut lainnya.[7] *Crawling*  membutuhkan data API Twitter atau Application Programming Interface (API) Twitter adalah suatu program atau aplikasi yang disediakan pihak twitter untuk memperoleh developer lain dalam mengakses informasi yang ada di website twitter.[8] Pada penelitian ini, penulis melakukan *crawling* menggunakan bahasa pemrograman *python* pada website *twitter.com* dengan kata kunci ‘Saham’. Data yang diperoleh berupa *tweet* dimana *tweet* yang diambil menggunakan bahasa inggris.

Setelah *crawling* selanjutnya adalah tahapan *preprocessing data,* dimana *preprocessing data* adalah sebuah metode yang digunakan untuk membersihkan data mentah untuk mempersiapkan proses prosedur lainnya.[9] Tujuan dari *preprocessing*  adalah mentransformasi data ke suatu format yang prosesnya lebih mudah dan efektif untuk kebutuhan peneiliti.[10] Data yang digunakan pada penelitian ini adalah tanggal *tweet* dibuat, teks *tweet,* nama pengguna *tweet¸* jumlah *favorite dan retweets* pada *tweet, device* yang digunakan oleh pengguna untuk melakukan *tweet* dengan kata kunci ‘Saham’.

*Cleaning data* digunakan untuk menbersihkan data dari informasi yang tidak berguna yang dapat memperlambat query ataupun memperburuk kualitas hasil dari data.[11] Pada tahap ini dikonfigurasi ulang sehingga mempunyai format yang konsisten.

Data yang telah di *preproccesing and cleaning* tersebut maka akan divisualisasikan dan diklasifikasikan.[12] Data-data tersebut akan divisualisasikan mulai kata terbanyak yang muncul pada *tweet* dengan kata kunci ‘Saham’, jumlah *tweet* berdasarkan waktu tertentu, jumlah *retweets* berdasarkan waktu tertentu, jumlah *favorite* berdasarkan waktu tertentu, jumlah *user status* pada waktu tertentu, jumlah *favorite* berdasarkam *device* yang digunakan, jumlah *retweets* berdasarkan *device,* jumlah *user status* berdasarkan *device,* jumlah *device* yang digunakan.

**III. PEMBAHASAN DAN HASIL**

Pada penelitian ini, penulis menggunakan bahasa *python* dengan *library tweepy, congpraser, pandas, re, itertools, collections, matplotlib.pyplot,* dan *seaborn.*[13] Pemilihan bahasa pemrograman tidak hanya dibatasi dengan bahasa *python*, pemilihan ini dilakukan dikarenakan bahasa pemrograman *python* dapat digunakan sebagai data analisis.[14]

Penelitian diawali dengan menentukan waktu, bahasa, dan jumlah data yang akan diambil dengan kata kunci Saham, dimana data yang diambil terdiri dari *time, id, user, name, user status count, tweet, favorite, retweets,* dan *device* sebanyak 500 data lalu dieksport ke file .csv dengan menggunakan source code sebagai berikut :

columns = ['Time', 'ID', 'User','Name','User Status Count', 'Tweet', 'Favorite','Retweets Count','Device']

data = []

key='saham'

for tweet in tweepy.Cursor(api.search, q=key, lang="en").items(500):

    data.append([tweet.created\_at, tweet.id, tweet.user.screen\_name, tweet.user.name, tweet.user.statuses\_count, tweet.text,

                 tweet.favorite\_count, tweet.retweet\_count,  tweet.source])

df = pd.DataFrame(data, columns = columns)

print(df)

df.to\_csv('RA-119140026-Saham.csv')

Tahap selanjutnya melakukan *preprocessing data* dengan melihat apakah ada data yang *missing value* atau *null* dengan *source code* sebagai berikut :

df.info()

Visualisasi pertama akan menampilkan banyaknya kata yang digunakan dalam *tweet* dengan kata kunci ‘Saham’, namun sebelum divisualisasikan perlu dilakukanya *cleaning data* dengan *source code* sebagai berikut:

def remove(txt):

  return re.sub("[^0-9A-za-z \t]|(\w+:\/\/\S+)","",txt)

tweet\_remove\_url = [remove(list\_tweet) for list\_tweet in all\_tweet]

tweet\_remove\_url[:10]

*Source code* di atas berfungsi untuk menghilangkan beberapa kata ataupun kalimat seperti www.,@, dan lain sebagainya supaya teks dari *tweet* bisa ditampilkan kata terbanyak dalam bentuk visual.

Visualisasi pertama akan menampilkan banyaknya kata yang digunakan dalam *tweet* dengan kata kunci ‘Saham’, namun sebelum divisualisasikan perlu dilakukanya *cleaning data* dengan *source code* sebagai berikut:

list\_tweet\_no\_url = list(iter.chain(\*word\_in\_tweet))

count\_word\_list\_tweet = collections.Counter(list\_tweet\_no\_url)

count\_word\_list\_tweet.most\_common(25)

*Source Code tersebut* berfungsi untuk mengsortir kata berdasarkan jumlah yang paling banyak muncul lalu divisualisasikan dengan *source code* sebagai berikut :

clean\_list\_tweet.sort\_values(by='Jumlah').plot.barh(x='Kata',y='Jumlah',color='green')

Maka hasil visualisasi dari jumlah kata pada *tweet* yang sering digunakan dengan kata kunci ‘Saham’ sebagai berikut :

Graphical user interface, application

Description automatically generated

Gambar 2. Visualisasi Kata Terbanyak

Berdasarkan informasi di atas dapat disumplkan bahwa kata ‘saham’ sering digunakan sebanyak 431 kali, kata ‘rt’ sebanyak 323 kali, kata ‘*to*’ digunakan sebanyak 316 kali, kata ‘*with*’ digunakan sebanyak 223 kali, dan kata ‘*never*’ digunakan sebanyak 196 kali pada *tweet* dengan kata kunci Saham.

Visualisasi kedua yaitu menampilkan informasi secara visual mengenai jumlah *tweet* pada jangka waktu tertentu. Seblum itu, perlu adanya *manipulating data* karena data yang kita ambil hanya tanggal tanpa menggunakan waktu dan menampilkan jumlah *tweet* dari waktu tertentu dengan *source code* sebagai berikut :

a=df['Time']

tanggal=a.dt.strftime('%Y-%m-%d')

b=tanggal.value\_counts()

b

Menampilkan hasil visualisai data yang sudah dimanipulating dengan bentuk diagram histogram, maka *source code* sebagai berikut :

plt.hist(tanggal,color='#4CAF50')

plt.xticks(rotation='vertical')

plt.legend(['Tweet'])

plt.xlabel('Tanggal')

plt.ylabel('Jumlah Tweet')

plt.show()

Graphical user interface, application

Description automatically generated

Gambar 3. Visualisasi Jumlah Tweet

Berdasarkan informasi di atas dapat dilihat bahwa *tweet* saham mengalami penurunan dari tanggal 22-03-22 ke 28-03-2022. Dimana pada tanggal 22-03-2022 terdapat sebanyak 120 *tweet,* pada tanggal 23-03-2022 terdapat sebanyak 88 *tweet*, pada tanggal 24-03-2022 terdapat sebanyak 82 *tweet,* pada tanggal 25-03-2022 terdapat sebanyak 76 *tweet,* pada tanggal 26-03-2022 terdapat sebanyak 54 *tweet,* pada tanggal 27-03-2022 terdapat sebanyak 47 *tweet,* dan pada tanggal 28-03-2022 terdapat sebanyak 33 *tweet.*

Visualisasi ketiga peneliti ingin menampilkan diagram garis dengan jumlah *retweets* berdasarkan jangka waktu tertentu dengan *source code* sebagai berikut :

plt.plot(df['Time'],df['Retweets Count'])

plt.xticks(rotation='vertical')

plt.legend(['Retweets'])

plt.xlabel('Waktu')

plt.ylabel('Jumlah Retweets')

plt.show()

Graphical user interface, application

Description automatically generated

Gambar 4. Visualisasi Jumlah Retweets dengan Tanggal

Berdasarkan visualisasi di atas jumlah *retweets* di atas mengalamim penurunan dan kenaikan secara signifikan. Pada jangka waktu 22-03-2022 sampai 23-03-2022 dengan waktu yang berbeda naik sampai 2000 lalu turun ke range 250 jumlah *tweets* dan pada jangka 25-03-2022 sampai 26-03-2022 menjadi jangka waktu dengan jumlah *retweets* paling sedikit.

Visualisas keempat peneliti ingin memvisualisasikan jumlah *user status* berdasarkan jangka waktu tertentu dengan *source code* dan tampilan sebagai berikut :

plt.plot(df['Time'],df['User Status Count'],color='hotpink')

plt.xticks(rotation='vertical')

plt.legend(['User Status'])

plt.xlabel('Waktu')

plt.ylabel('Jumlah User Status')

plt.show()

Graphical user interface, text, application

Description automatically generated

Gambar 5. Visualisasi Jumlah User Status pada Tanggal

Berdasarkan visualisasi di atas dapat disimpulkan bahwa informasi yang didapat adalah pada jangka waktu 24-03-2022 jumlah *user status* terbanyak menyentuh tanggal 390000 kali, pada jangka tanggal 22-03-2022 sebanyak 300000 kali jumlah *user status,* dan pada jangka tanggal 26-03-2022 sebanyak 250000 jumlah *user status*.

Visualisasi kelima yaitu mevisualisasikan jumlah *like* pada jangka waktu tertentu dengan *source code* dan tampilan sebagai berikut:

plt.plot(df['Time'],df['Favorite'],color='red')

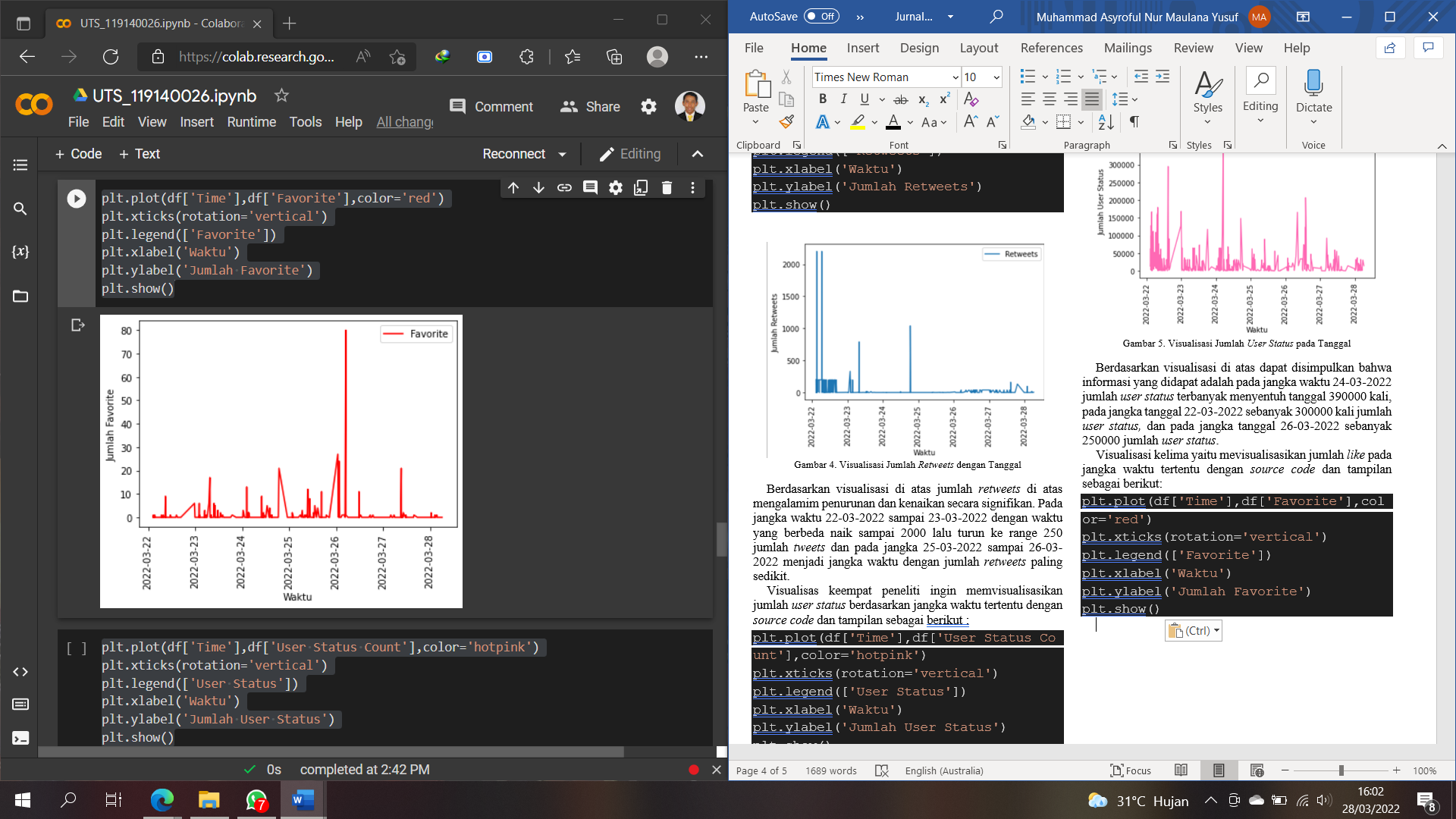
plt.xticks(rotation='vertical')

plt.legend(['Favorite'])

plt.xlabel('Waktu')

plt.ylabel('Jumlah Favorite')

plt.show()



Gambar 6. Visualisasi Jumlah Like Terhadap Tanggal

Berdasarkan visualisasi di atas maka dapat dismpulkan pada jangka waktu 26-03-2022 menjadi jangka waktu terbanyak like pada *tweet* dengan kata kunci ‘Saham’ dan pada jangka waktu 27-03-2022 dan 28-03-2022 menjadi jangka wakti dengan jumlah *like* sedikit pada *tweet* dengan kata kunci ‘Saham’.

Visualisasi keenam peniliti ingin memvisualisasikan jumlah *like* berdasarkan *device* yang digunakan pengguna ketika melakukan *tweet* dengan kata kunci ‘Saham’. Berikut *source code* dan tampilan dari visualisasi tersebut:

plt.bar(df['Device'],df['Favorite'],color='#4CAF50')

plt.xticks(rotation='vertical')

plt.legend(['Like'])

plt.xlabel('Device')

plt.ylabel('Jumlah Like')

plt.show()

Graphical user interface

Description automatically generated

Gambar 7. Jumlah Like Terhadap Device

Berdasarkan visualisasi di atas dapat disimpulkan bahwa Twitter Web App menjadi *device* yang banyak digunakan untuk *like* pada *tweet* dengan kata kunci ‘Saham’ disusul Twitter for iPhone dan Twitter for Android. Pengguna lebih senang menggunakan Twitter Web App ketika melakukan *like* pada *tweet* saham.

Visualisasi ketujuh peniliti ingin memvisualisasikan jumlah *retweets* pada *tweet* saham terhadap *device* yang digunakan oleh pengguna. Berikut *source code* dan tampilan visualisasi tersebut:

plt.bar(df['Device'],df['Retweets Count'],color='hotpink')

plt.xticks(rotation='vertical')

plt.legend(['Retweets'])

plt.xlabel('Device')

plt.ylabel('Jumlah Retweets')

plt.show()

Graphical user interface

Description automatically generated

Gambar 8. Visualisasi Jumlah Retweets Terhadap Device

Berdasarkan visualisasi di atas dapat disimpulkan bahwa pengguna lebih menyukai menggunakan Twitter for Android dan Twitter for iPhone dalam melakukan *retweets* pada *tweet* dengan kata kunci Saham.

Visualisasi kedelapan peneliti ingin memvisualisasikan jumlah *user status* terharap *device* yang digunakan oleh pengguna di twitter.com dengan *source code* dan tampilan sebagai berikut :

plt.bar(df['Device'],df['User Status Count'])

plt.xticks(rotation='vertical')

plt.legend(['User Status'])

plt.xlabel('Device')

plt.ylabel('Jumlah User Status')

plt.show()

Graphical user interface

Description automatically generated

Gambar 9. Jumlah User Status Terhadap Device

Berdasarkan visualisasi di atas maka dapat disimpulkan bahwa jumlah *user status* banyak dilakukan oleh pengguna menggunakan *device* Twitter for Android, Twitter for iPhone, dan Twitter Web App. Terdapat sedikit perbedaan ketika melihat visualisasi-visualisasi sebelumnya bahwa pengguna memiliki jumlah *device* yang berbeda pada jumlah *like, retweets,* dan *user status*.[15]

Visualisasi selanjutnya adalah peniliti ingin memvisualisasikan jumlah *device* terhadap *tweet* dengan kata kunci saham. Berikut *source code* dan tampilan visualisasi sebagai berikut :

plt.hist(df['Device'],color='#4CAF50')

plt.xticks(rotation='vertical')

plt.legend(['Device'])

plt.xlabel('Histogram Device')

plt.ylabel('Jumlah Device')

plt.show()

Graphical user interface, application

Description automatically generated

Gambar 10. Jumlah Device Terhadap Tweet Saham

Berdasarkan visualisasi di atas pengguna melakuka *tweet* dengan kata kunci saham lebih banyak menggunakan *device* Twitter for Android dan Twitter for iPhone.

**IV. Kesimpulan**

Berdasarkan dari penelitian yang telah dilakukan, dapat disimpulkan bahwa *tweet* dengan kata kunci saham memiliki daya tarik tersendiri bagi pengguna Twitter untuk melakukan *tweet, like, retweets,* dan *user status* berdasarkan jangka waktu tertentu dan *device* yang digunakan oleh pengguna.

**Daftar Pustaka**

[1] H. Bakshi, “Framework for crawling and local event detection using twitter data,” 2011, Accessed: Mar. 28, 2022. [Online]. Available: https://search.proquest.com/openview/f2854cc38efaca739725cc29c51ef22f/1?pq-origsite=gscholar&cbl=18750

[2] E. Rusliati, E. F.-J. B. dan Akuntansi, and undefined 2010, “Pemecahan saham terhadap likuiditas dan return saham,” *jurnaltsm.id*, Accessed: Mar. 28, 2022. [Online]. Available: http://jurnaltsm.id/index.php/JBA/article/view/213

[3] D. Arista, A. A.-J. I. M. dan Akuntansi, and undefined 2012, “Analisis Faktor–Faktor Yang Mempengaruhi Return Saham,” *jurnal.stietotalwin.ac.id*, Accessed: Mar. 28, 2022. [Online]. Available: http://jurnal.stietotalwin.ac.id/index.php/jimat/article/download/29/28

[4] S. J. Teller and C. H. Séquin, “Visibility preprocessing for interactive walkthroughs,” *ACM SIGGRAPH Computer Graphics*, vol. 25, no. 4, pp. 61–70, Jul. 1991, doi: 10.1145/127719.122725.

[5] Y. Tresnawati, “SENTIMENT ANALYSIS OF TWITTER USING AGGLOMERATIVE HIERARCHICAL CLUSTERING A THESIS Presented as Partial Fulfillment of Requirements to Obtain Sarjana Komputer Degree in Informatics Engineering Department.”

[6] A. Famili, W. Shen, … R. W.-I. data, and undefined 1997, “Data preprocessing and intelligent data analysis,” *content.iospress.com*, Accessed: Mar. 28, 2022. [Online]. Available: https://content.iospress.com/articles/intelligent-data-analysis/ida1-1-02

[7] M. Bošnjak, E. Oliveira, … J. M.-P. of the 21st, and undefined 2012, “Twitterecho: a distributed focused crawler to support open research with twitter data,” *dl.acm.org*, Accessed: Mar. 28, 2022. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/2187980.2188266

[8] S. Sohail, M. Khan, M. Arsalan, … A. K. preprint arXiv, and undefined 2021, “Crawling Twitter data through API: A technical/legal perspective,” *arxiv.org*, Accessed: Mar. 28, 2022. [Online]. Available: https://arxiv.org/abs/2105.10724

[9] A. Dong *et al.*, “Time is of the essence: Improving recency ranking using Twitter data,” *Proceedings of the 19th International Conference on World Wide Web, WWW ’10*, pp. 331–340, 2010, doi: 10.1145/1772690.1772725.

[10] M. Bošnjak, E. Oliveira, J. Martins, E. M. Rodrigues, and L. Sarmento, “TwitterEcho - A distributed focused crawler to support open research with twitter data,” *WWW’12 - Proceedings of the 21st Annual Conference on World Wide Web Companion*, pp. 1233–1239, 2012, doi: 10.1145/2187980.2188266.

[11] R. McCreadie, I. Soboroff, J. Lin, C. Macdonald, I. Ounis, and D. McCullough, “On building a reusable Twitter corpus,” *SIGIR’12 - Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1113–1114, 2012, doi: 10.1145/2348283.2348495.

[12] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurahman Baizal, “Data Crawling Otomatis pada Twitter,” Sep. 2016, pp. 11–16. doi: 10.21108/indosc.2016.111.

[13] J. R.-O. http://tweepy. readthedocs. io/en/v3 and undefined 2009, “tweepy Documentation,” *media.readthedocs.org*, 2018, Accessed: Mar. 28, 2022. [Online]. Available: https://media.readthedocs.org/pdf/tweepy/v3.6.0/tweepy.pdf

[14] E. Negara, R. Andryani, P. S.-I. Journal, and undefined 2016, “Analisis data twitter: Ekstraksi dan analisis data g eospasial,” *jurnal.informatika.lipi.go.id*, Accessed: Mar. 28, 2022. [Online]. Available: http://jurnal.informatika.lipi.go.id/index.php/inkom/article/view/433

[15] P. Meilina, “PENERAPAN DATA MINING DENGAN METODE KALSIFIKASI MENGGUNAKAN DECISION TREE DAN REGRESI,” 2015.