

Final Project: Explorasi dan Visualisasi Data

Visualisasi Data Teks pada Topik Formula E di Twitter

Student ID	Student name	Contribution description	Contribution (%)
2106638526	Aurelio Naufal Effendy	Mengambil data dari twitter, mengerjakan preprocessing data, dan membuat laporan akhir	100%
2106725103	Christian Audy Samuel Tampubolon	Melakukan pengolahan dan analisis data dan membantu membuat laporan akhir	100%
2106706205	Favian Sulthan Wafi	Melakukan pengolahan dan analisis data serta membuat laporan akhir	100%
2106638204	Rifqi Hafizuddin	Melakukan preprocessing data dan membuat laporan akhir	100%

Bagian 1. Pendahuluan

Pada hari Sabtu (4/6/2022) kemarin, ajang balapan Formula E telah digelar di Indonesia untuk pertama kali dalam sejarah, tepatnya di Jakarta International E-Prix Circuit (JIEC), Ancol, Jakarta Utara. Acara ini diperkirakan menyedot kurang lebih 60 ribu penonton. Ajang lomba balap mobil ramah lingkungan yang digagas oleh Gubernur DKI Jakarta, Anies Baswedan ini juga dihadiri para pejabat-pejabat Republik Indonesia, tak terkecuali Presiden Jokowi. Acara ini kerap menimbulkan pro dan kontra di kalangan masyarakat.

Oleh karena itu, kami ingin mencari tahu bagaimana tanggapan masyarakat terhadap berlangsungnya acara ini dengan melihat dan menganalisis cuitan masyarakat di aplikasi media sosial twitter. Kami juga ingin melihat kata-kata apa saja yang sering muncul terkait dengan topik Formula E di kalangan masyarakat.

Sumber data yang digunakan adalah data dari twitter (crawling data twitter menggunakan twitter API) dengan ukuran data sebesar 31 mb atau sebanyak 5000 tweets.

Bagian 2. Pre-processing

Langkah-langkah yang dilakukan dalam proses preprocessing sebagai berikut.

1. Pertama-tama, import module-module yang diperlukan, seperti module `tau-data`, `tweepy`, `tqdm`, `textblob`, dan lain-lain (untuk lebih jelasnya ada di file jupyter notebooknya)
2. Lalu, inisiasikan variabel untuk menaruh key access dari twitter API yang selanjutnya akan digunakan untuk login
3. Setelah login dengan API key barusan, inisiasikan query yang diisi dengan topik yang akan dicari yakni formula e. Karena akan digunakan Bahasa Indonesia untuk mencari tweetnya, otomatis semua keyword formula e pasti membahas formula e jakarta yang baru saja dilaksanakan
4. Set jumlah tweetnya sebanyak 5000 tweet lalu mulai data crawling dari twitter menggunakan query tadi, bahasa Indonesia, dan tweet modenya extended (supaya menampilkan keseluruhan tweet).
5. Setelah itu masukan hasil crawling data tersebut dalam suatu variabel dalam bentuk file json.
6. Buat fungsi untuk menyimpan hasil crawling twitter secara utuh, dan juga fungsi untuk menload kembali data yang sudah di crawling barusan. Sehingga data dapat diload kembali jika analisis ingin dilakukan di lain waktu.
7. Simpan hasil crawling twitter dalam file json dengan fungsi tadi. Setelah itu , data dapat diload kembali dan disimpan dalam variabel baru untuk selanjutnya ke proses cleaning data.
8. Inisiasikan variabel baru yang berisi tweet saja dari hasil crawling twitter yang telah diperoleh untuk digunakan dalam proses cleaning data.
9. Dalam proses cleaning data, yang pertama dilakukan adalah menghapus link url pada tweet. Import module `re` kemudian gunakan fungsi `compile` dan `sub` untuk menghapus link url pada tweet (link berbentuk `t.co`). Ubah bentuk `'tweets_saja'` dari list menjadi string supaya dapat dibersihkan.
10. Gunakan module dan fungsi yang sama seperti step 9 untuk menghapus username pada tweet.
11. Gunakan module dan fungsi yang sama seperti step 9 untuk mengubah hashtags dalam tweet menjadi tulisan.
12. Hapus non alpha-numeric (terutama emote) dengan fungsi `sub`.
13. Kemudian akan dilakukan tokenisasi dan penanganan slang atau singkatan di data teks. Tambahkan beberapa slang baru yang sering muncul pada tweets.
14. Lakukan filtering atau menggunakan Stopwords untuk menghilangkan kata-kata yang sering muncul dengan module `textblob`.
15. Simpan hasil preprocessing dengan bentuk file `txt`.

https://colab.research.google.com/drive/1AXeC0Np2pnmOdrvhjx4bhl7s_FKnkMX#scrollTo=Mk0HNf62NCcA

Bagian 3. Pengolahan dan Analisis Data

1. Word Cloud

Menggunakan Word Cloud dengan 55 terms dari 53,445 kata dan 3,928 kata yang unik, terlihat bahwa kata-kata yang paling sering muncul di data adalah: sukses(2082), anies(1087), islam(862), front(859), dan pembela(859). Terdapat kata-kata “front”, “pembela”, dan “islam” dengan jumlah yang hampir sama yang kelihatannya merupakan kepanjangan dari “FPI” yang seharusnya bisa dituliskan dengan satu kata saja. Dengan demikian, agar tidak memengaruhi Word Cloud, kata “front” dan “pembela” akan menjadi StopWords dan diwakilkan oleh kata “islam”.

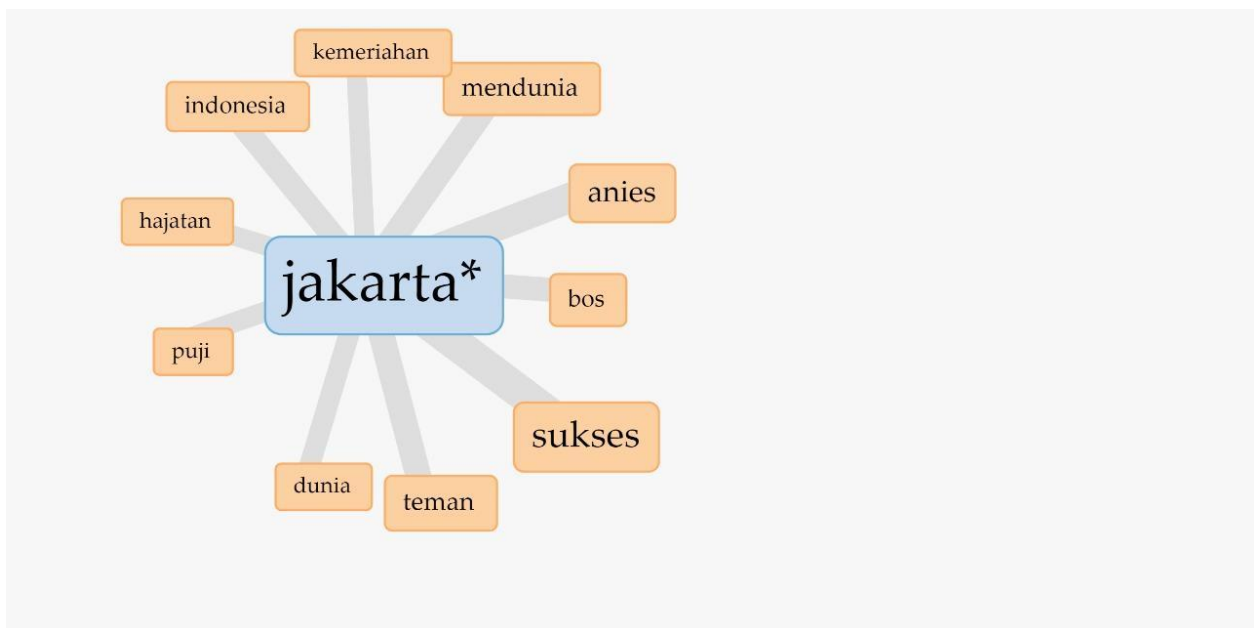


Kata-kata yang paling sering muncul menjadi: sukses(2082), anies(1087), islam(862), gubernur(709), dan jakarta (657).

2. Word Links

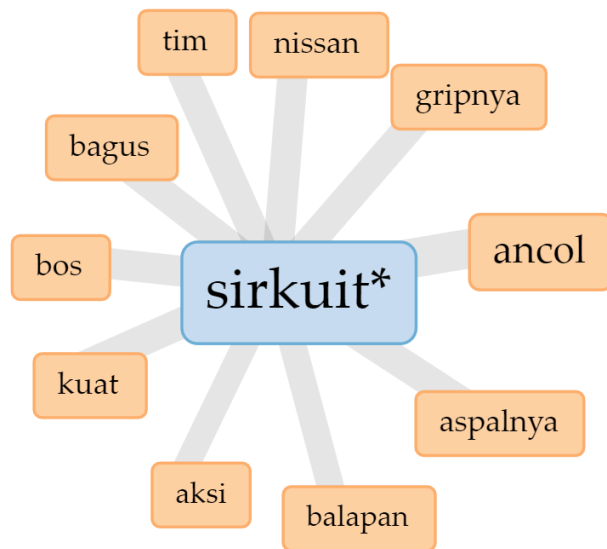
Selanjutnya akan digunakan Word Links untuk mengetahui hubungan antar kata-kata.

Kata utama: jakarta



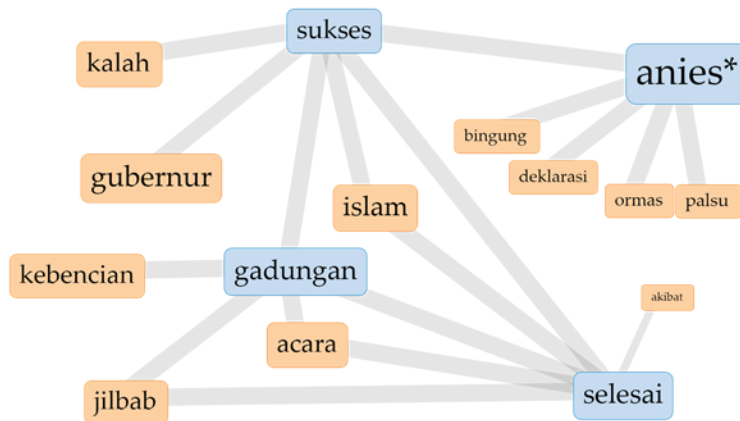
Terlihat bawah kata “jakarta” yang merupakan tempat dilaksanakannya Formula E berhubungan dengan kata-kata “puji”, “hajatan”, “sukses”, “mendunia”, dan “kemeraian”. Hal ini menandakan banyaknya tanggapan positif dari orang-orang terhadap penyelenggaraan Formula E di Jakarta.

Kata utama: sirkuit



Terlihat bawah kata “sirkuit” berhubungan dengan kata-kata “gripnya” dan “aspalnya” serta beberapa kata yang positif, seperti “bagus” dan “kuat”. Hal ini menandakan bahwa saat membicarakan “sirkuit” Formula E, banyak orang yang memberikan tanggapan positif dan menanggapi kualitas dari grip dan aspalnya. Kemudian ada kata “bos” yang cukup banyak diperbincangkan karena baru-baru ini bos Formula E memberikan apresiasi kepada gubernur DKI Jakarta atas keberhasilan penyelenggaraan Formula E dari segala sisi, termasuk sirkuit.

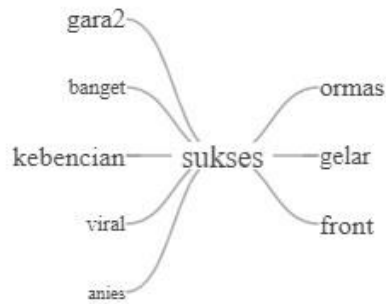
Kata utama : anies



Terlihat bawah kata “anies” yang merupakan gubernur DKI Jakarta sekaligus penyelenggara Formula E berhubungan dengan kata yang positif, yaitu sukses. Setelah ditelusuri lebih lanjut, kata “sukses” di sini berhubungan dengan kata “selesai” dan “gubernur”. Kemudian kata “selesai” di sini berhubungan dengan kata “acara” yang menandakan saat orang membicarakan “anies” banyak yang mengaitkan dengan kata “sukses” sebagai gubernur dan dalam penyelenggaraan acara Formula E sampai selesai. Akan tetapi, terdapat juga kata yang negatif seperti “gadungan” yang ternyata berhubungan juga dengan kata “acara” dan “sukses” yang menandakan bahwa saat membicarakan “anies” terdapat pula orang yang mengaitkan kata “gadungan” dalam konteks kesuksesan acara Formula E. Kemudian, terdapat pula kata-kata yang cukup politis saat membicarakan “anies”, yaitu “presiden” dan “2024”.

3. Word Tree

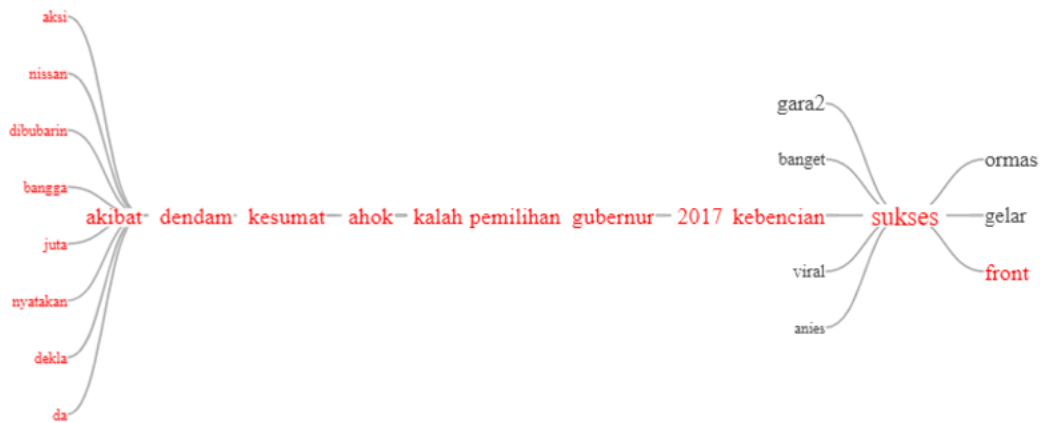
Setelah dilihat hubungan antarkatanya, agar lebih jelas akan digunakan Word Tree untuk melihat urutan atau frasa dari suatu kata.



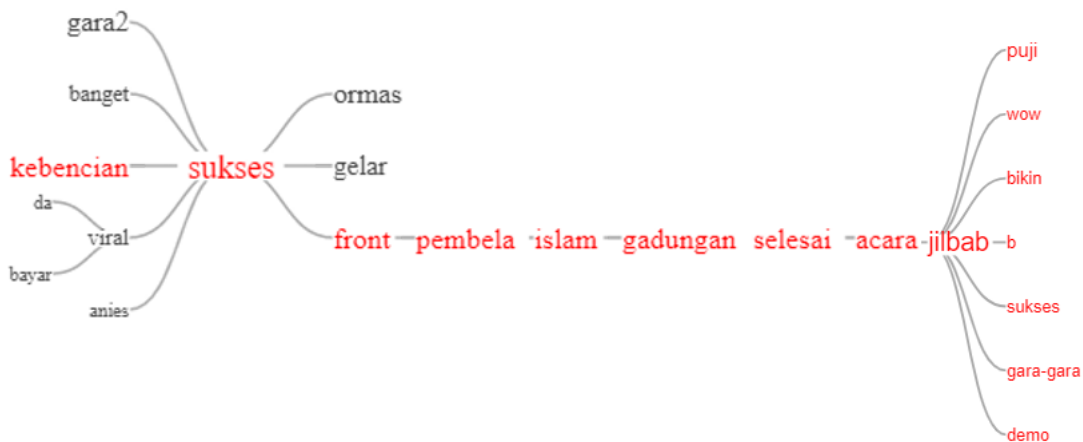
Dari Word Tree tersebut diketahui kata “sukses” ada banyak hubungan dengan kata sebagai berikut.



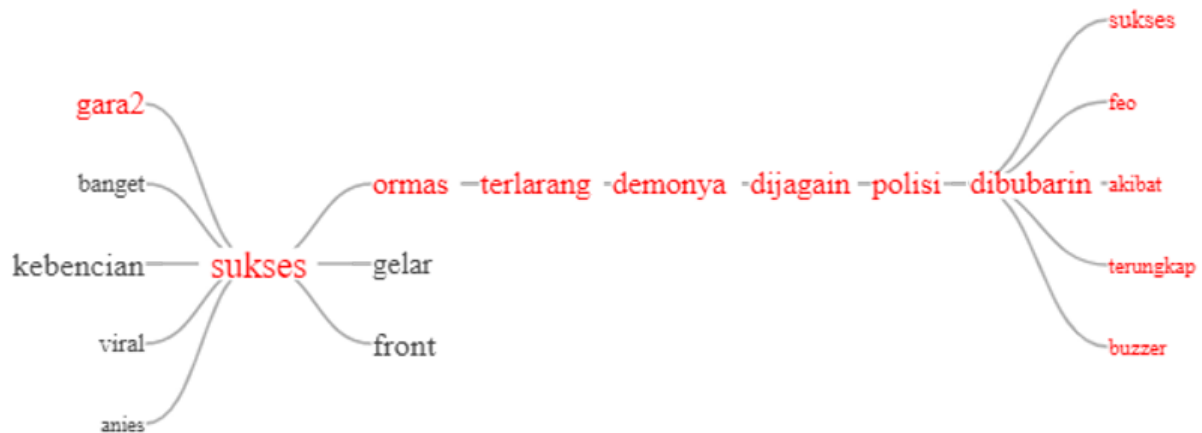
Setelah diexpand kata “gara2” berhubungan dengan kata “stress” dan kata “stress” juga berhubungan dengan kata “genjer”. Bisa diperkirakan bahwa beberapa orang beropini bahwa “formula e sukses gara2 stress genjer”. Setelah kata “gelar” diexpand kata “gelar” berhubungan dengan kata “event”, “hajatan”, “terbesar”. Bisa diperkirakan sumber data beropini “sukses formula e itu gelar event, gelar hajatan, atau gelar terbesar”.



Setelah kata “kebencian” diekspand kata-kata sebagai berikut ditampilkan. Bisa diperkirakan menurut beberapa orang Gubernur Ahok benci dengan dengan sukses formula e dan akibatnya berhubungan dengan kata “aksi”, “Nissan”, “dibubarin”, dst.



Setelah kata “viral” diekspand ditampilkan kata “da” dan “bayar”, diperkirakan menurut beberapa orang “sukses formula e viral karena dibayar”. Dan setelah kata front diekspand kata-kata sebagai berikut ditampilkan. Diperkirakan menurut beberapa orang “frontnya sukses formula e adalah pembela islam yang gadungan dengan acara-acara puji, demo, dst”.



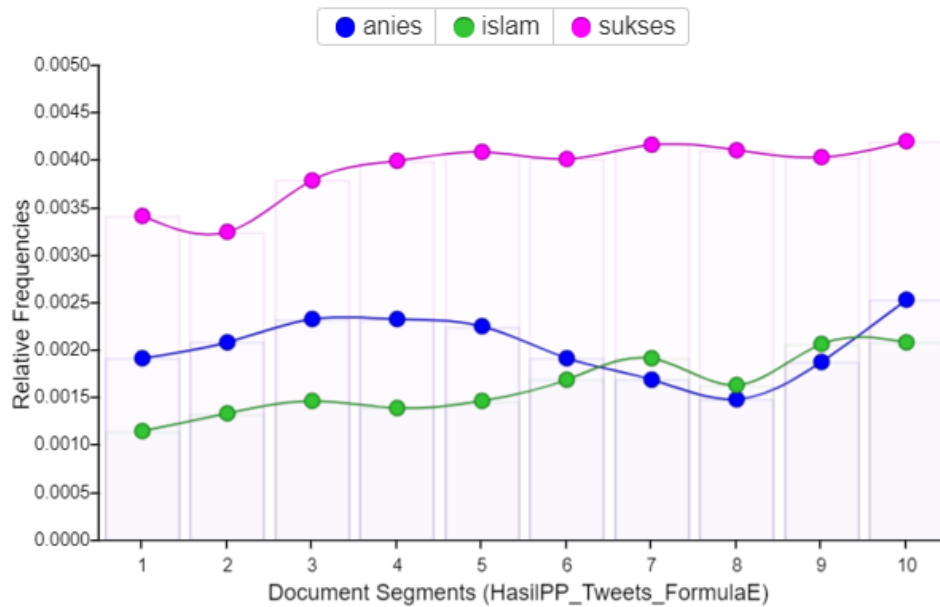
setelah kata “ormas” diekspand kata-kata sebagai beriku ditampilkan. Bisa diperkirakan beberapa sumber data memperkirakan “ormas terlarang melakukan demo dan ormas dijagain dan dibubarin oleh polisi”.



Setelah kata “anies” diexpand kata-kata sebagai berikut ditampilkan. Diketahui bahwa ada banyak sumber data yang beropini tentang anies, hanya diketahui opini tersebut berhubungan dengan kata-kata “formul”, “acara”, “iklan”, dst.

4. Trends

Terakhir, kami menggunakan Trends untuk melihat banyaknya penggunaan suatu kata.



Setelah menggunakan tools Trends ditampilkan diagram ini. Dari diagram ini diketahui kata-kata atau opini-opini yang paling banyak dari sumber data yang digunakan. Dari diagram tersebut diketahui kata “sukses” paling sering digunakan, jadi bisa diperkirakan banyak orang beropini “formula e sukses”. Dari diagram tersebut diketahui kata “anies” dan “islam” lumayan banyak digunakan saat berbicara tentang formula e.

Bagian 4. Penutup

Berdasarkan hasil preprocessing dan analisis visualisasi data tweet dengan topik "Formula E" yang kami lakukan, tanggapan dari masyarakat atas terlaksananya ajang Formula E di Jakarta mayoritas positif. Banyak masyarakat yang setuju Formula E adalah acara yang sukses. Banyak juga yang memberikan apresiasi terhadap para penyelenggara Formula E atas terlaksananya acara dengan sukses sampai selesai. Anies Baswedan selaku Gubernur DKI Jakarta juga mendapatkan respons positif dari masyarakat. Banyak yang beropini bahwa kesuksesan ajang Formula E adalah kesuksesan untuk Anies sebagai gubernur. Akan tetapi, terdapat juga beberapa tanggapan negatif dari masyarakat. Tanggapan negatif yang paling banyak muncul adalah kesuksesan ajang Formula E dan respons positif terhadap Anies Baswedan hanyalah gadungan atau dibuat-buat saja oleh FPI. Dapat disimpulkan juga, acara Formula E ini sering dikaitkan dengan politik, terlihat dari kata-kata yang sering muncul dalam visualisasi.

Link Google Drive untuk folder terkompresi:

<https://drive.google.com/drive/folders/1LHfE6JWSc1VgiYATmQkAPsse4O580i00?usp=sharing>

(Sumber data diambil dari twitter)

Link video presentasi singkat :

<https://www.youtube.com/watch?v=M1ndBqh6VHU>