

**LAPORAN PROYEK AKHIR
PRAKTIKUM DATA SCIENCE
TWEET SEPUTAR TED CRUZ DAN AOC**



RISKY AULIA NUGRAHA
AFIFRENDRA RIFQI NUGRAHA

123190002
123190012

**PROGRAM STUDI INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK INDUSTRI
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
YOGYAKARTA
2021**

1. PENDAHULUAN

Pada era teknologi yang semakin meningkat dan diikuti oleh perkembangan TI yang semakin modern berpengaruh terhadap kehidupan individu maupun kelompok. Perkembangan TI dapat membantu dalam mengembangkan tugas-tugas baru pada perusahaan berskala pasar global atau pada instansi pemerintah, TI dimanfaatkan untuk mengatasi meningkatkan kemampuan seseorang dalam meraih keunggulan dan kesuksesan yang handal serta canggih.

Pemanfaatan teknologi informasi juga digunakan dalam instansi pemerintahan yang dapat digunakan disegala aspek dengan menggunakan data yang telah tersedia, data ini akan diolah dengan teknologi yang akan menghasilkan informasi transparan, akurat dan mudah dipahami oleh semua kalangan.

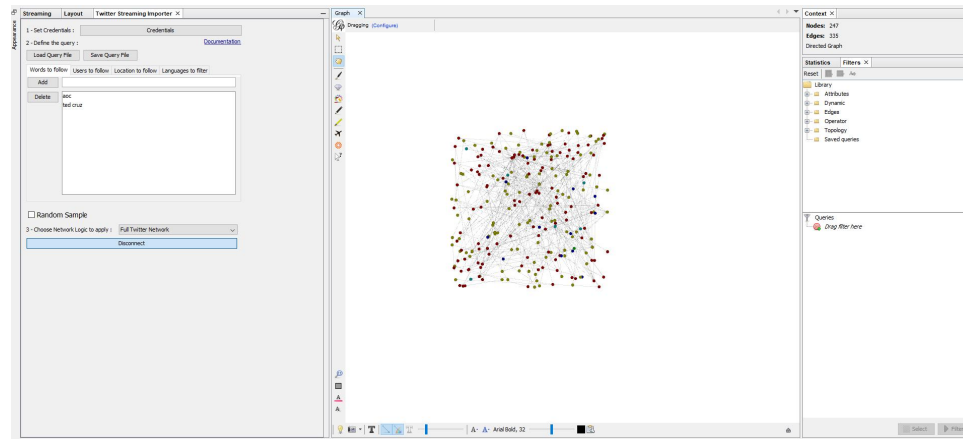
Salah satu isu pemerintahan yang sedang ramai diperbincangan masyarakat adalah pemerintahan di Amerika Serikat, Ted Cruz dan Alexandria Ocasio-Cortez (AOC) yang membahas banyak hal di masing masing akun twiter yang mereka punya. Tweet yang mereka upload di twiter ini menimbulkan banyak pandangan dari masyarakat seluruh dunia dan pandangan ini menjadi data yang bisa dianalisis untuk mengetahui bagaimana opini masyarakat terhadap tweet tersebut.

Analisis tersebut biasa disebut sebagai Analisa Sentimen atau *Sentiment Analysis* ini adalah proses penggunaan text analisis untuk mendapatkan berbagai sumber data dari internet dan beragam platform media sosial. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada platform tersebut.

Dengan menggunakan teknologi dan sekumpulan data yang ada, dapat menghasilkan informasi yang akurat juga mudah dipahami tanpa perlu waktu yang lama. Menggunakan metode KNN untuk mengolah datanya dan Bahasa R yang digunakan untuk membuat aplikasinya.

2. METODE

Data yang digunakan didapat dari tweet masyarakat seluruh dunia yang berada di media sosial, Twitter. Menggunakan aplikasi Gephi disertai dengan plugin bernama Twitter Streaming Importer. Dengan membuat akun twitter developer untuk mendapatkan twitter API yang akan dihubungkan ke dalam aplikasi Gephi sehingga memungkinkan untuk mendapatkan data dari twitter.



The screenshot shows the Gephi Twitter Streaming Importer interface with the 'Data Table' tab selected. The table displays a list of tweets with columns for 'id', 'label', 'timestamp', 'text', 'lang', 'place_name', 'place_type', 'place_subtype', 'place_subtype', 'created_at', 'lang', 'possibly_sensitive', 'quoted_status_permalink', 'description', 'email', 'profile_image_url', 'friends_count', 'followers_count', and 'real_name'. The table contains 247 rows of data, each representing a tweet. The 'id' column contains unique identifiers for each tweet, and the 'text' column contains the tweet content. The 'timestamp' column shows the date and time of the tweet. The 'lang' column indicates the language of the tweet. The 'place_name' column shows the location of the tweet. The 'place_type' column shows the type of location. The 'place_subtype' column shows the subtype of location. The 'created_at' column shows the date and time of the tweet. The 'lang' column shows the language of the tweet. The 'possibly_sensitive' column shows whether the tweet is possibly sensitive. The 'quoted_status_permalink' column shows the permalink of the tweet. The 'description' column shows the description of the tweet. The 'email' column shows the email address of the user. The 'profile_image_url' column shows the profile image URL of the user. The 'friends_count' column shows the number of friends the user has. The 'followers_count' column shows the number of followers the user has. The 'real_name' column shows the real name of the user.

Metode yang digunakan berupa metode KNN (K-Nearest Neighbor) merupakan salah satu metode untuk mengambil keputusan menggunakan pembelajaran terawasi dimana hasil dari data masukan yang baru diklasifikasi berdasarkan terdekat dalam data nilai.

Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut. KNN merupakan algoritma supervised learning dimana hasil dari query instance yang baru diklasifikasi berdasarkan mayoritas dari kategori pada algoritma KNN. Dimana kelas yang paling banyak muncul yang nantinya akan menjadi kelas hasil dari klasifikasi.

Kedekatan didefinisikan dalam jarak metrik, seperti jarak Euclidean. Jarak Euclidean dapat dicari dengan menggunakan persamaan 1 berikut ini:

$$D_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan :

D : jarak kedekatan

x : data training

y : data testing

n : jumlah atribut individu antara 1 s.d. n

f : fungsi similarity atribut antara kasus dan kasus

i = Atribut individu antara 1 sampai dengan n

Langkah-langkah untuk menghitung metode K

Nearest Neighbor, antara lain :

- : 1. Menentukan parameter (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak Euclid (query instance) masing-masing objek terhadap data sampel yang diberikan menggunakan persamaan 1.
3. Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak Euclid terkecil.
4. Mengumpulkan kategori (Klasifikasi Nearest Neighbor)
5. Dengan menggunakan kategori Nearest Neighbor yang paling mayoritas maka dapat diprediksi nilai query instance yang telah dihitung.

```
# knn
k <- 5

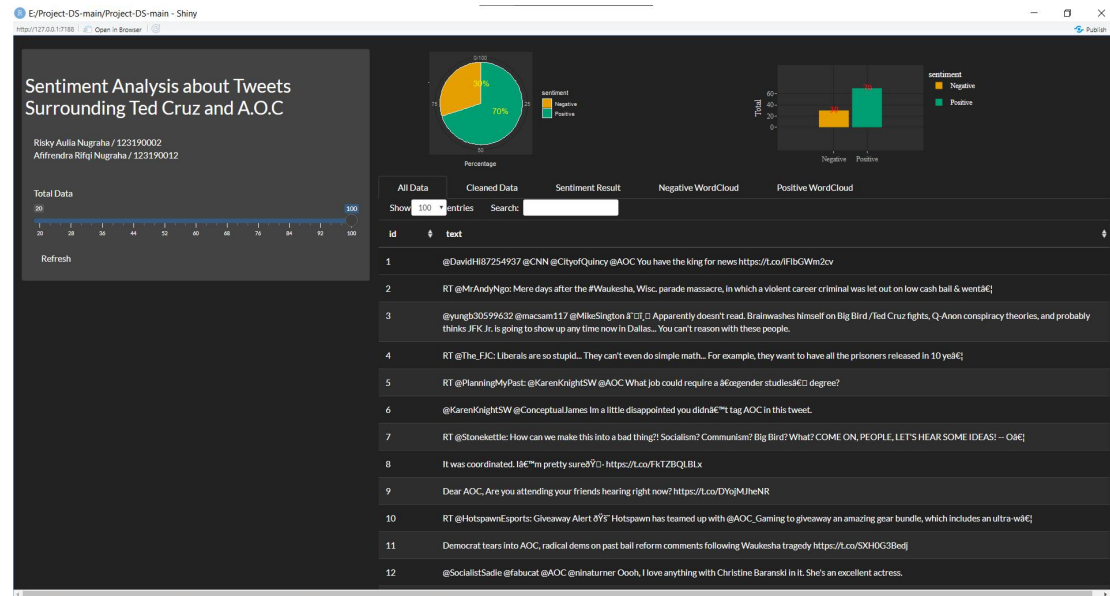
cek <- cosine %>%
  left_join(train_data, by = "id") %>%
  select(id, cosine, sentiment) %>%
  arrange(desc(cosine)) %>%
  head(k)

predict_sentiment <- cek %>%
  count(sentiment)
predict_sentiment <- predict_sentiment$sentiment[which.max(predict_sentiment$n)]

predict_result$sentiment[j] <- predict_sentiment
}
write.csv(predict_data_all, file = "data-raw/predict_data_all.csv", row.names = FALSE)
write.csv(predict_data_clean, file = "data-raw/predict_data_clean.csv", row.names = FALSE)
write.csv(predict_result, file = "data-raw/predict_result.csv", row.names = FALSE)
cat(sprintf("\nselesai"))
```

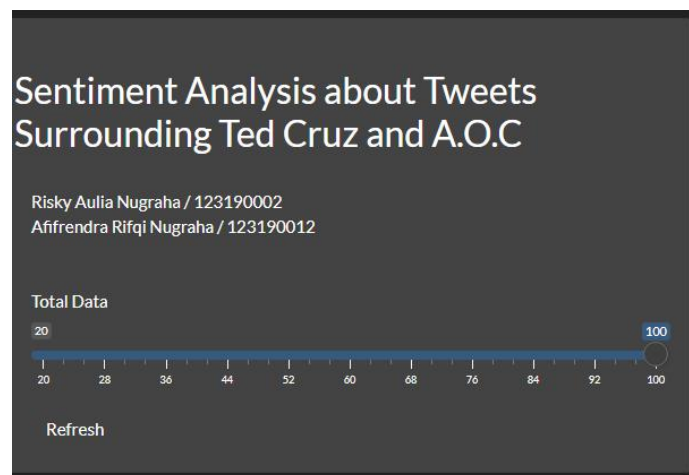
Code pada R studio yang menggunakan metode KNN untuk menentukan nilai sentiment analisis pada tweet tersebut. Ini bertujuan untuk memberikan informasi tweet positive atau negative.

3. HASIL DAN PEMBAHASAN



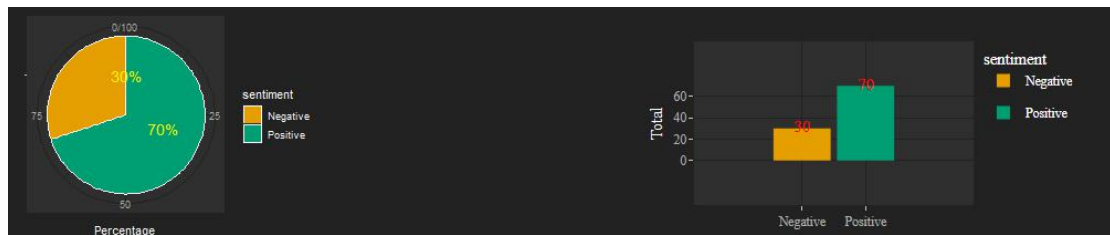
Gambar 3.1

Pada gambar 3.1 merupakan halaman awal ketika program dijalankan, terdapat berbagai fitur yang bisa digunakan untuk mendapatkan informasi secara lengkap dan mudah dipahami.



Gambar 3.2

Gambar 3.2 menampilkan judul program dan penggunaan total data tweet yang ingin ditampilkan pada program, total data ini bisa disesuaikan oleh keinginan user, parameter data dari 20 sampai 100, user hanya tinggal menarik lingkaran kearah kanan atau kiri menyesuaikan berapa banyak data yang ingin ditampilkan.



Gambar 3.3

Gambar 3.3 ini berupa keterangan dengan bentuk diagram batang dan diagram lingkaran, menjelaskan presentasi tweet positive dan negatif secara sederhana.



Gambar 3.4

Gambar 3.4 ini memperlihatkan fitur lain yang berada diatas diagram batang, fitur ini muncul otomatis karna menggunakan library plotly yang mana memang plotly sendiri menyediakan tools untuk mendukung interaksi pada diagram batang (plot). Fitur/tools dari plotly ini berupa Export to png, Zoom, Selection dan Hovering.

Processing...					
All Data		Cleaned Data	Sentiment Result	Negative WordCloud	Positive WordCloud
Show	10 entries	Search: <input type="text"/>			
	10				
	25				
	50				
	100				
id	text				
1	@DavidH87254937 @CNN @CityofQuincy @AOC You have the king for news https://t.co/IFibGWm2cv				
2	RT @MrAndyNgo: Mere days after the #Waukesha, Wisc. parade massacre, in which a violent career criminal was let out on low cash bail & wentâ€¦				
3	@yungb30599632 @macsam117 @MikeSington â€œ Apparently doesn't read. Brainwashes himself on Big Bird /Ted Cruz fights, Q-Anon conspiracy theories, and probably thinks JFK Jr. is going to show up any time now in Dallas... You can't reason with these people.				
4	RT @The_FJC: Liberals are so stupid... They can't even do simple math... For example, they want to have all the prisoners released in 10 yeâ€¦				
5	RT @PlanningMyPast: @KarenKnightSW @AOC What job could require a â€œgender studiesâ€œ degree?				
6	@KarenKnightSW @ConceptualJames Im a little disappointed you didnâ€™t tag AOC in this tweet.				
7	RT @Stonekettle: How can we make this into a bad thing?! Socialism? Communism? Big Bird? What? COME ON, PEOPLE, LET'S HEAR SOME IDEAS! -- Oâ€¦				
8	It was coordinated. Iâ€™m pretty sureâ€¦ https://t.co/FK1ZBQLBLx				
9	Dear AOC, Are you attending your friends hearing right now? https://t.co/DYojMJheNR				
10	RT @HotspawnEsports: Giveaway Alert â€œ Hotspawn has teamed up with @AOC_Gaming to giveaway an amazing gear bundle, which includes an ultra-wâ€¦				
id	text				
Showing 1 to 10 of 100 entries Previous 12345...10 Next					

Gambar 3.5

Pada Gambar 3.5 ini menampilkan bagian All Data, data mentah dari twitter yang belum siap untuk diinformasikan. Disetiap bagian (All Data, Cleaned Data, Sentiment Result, Negative WordCloud dan Positive Cloud) dapat memilih berapa banyak data yang ingin ditampilkan fitur ini bernama show, dan terdapat search untuk mencari tweet yang diinginkan.

All Data		Cleaned Data	Sentiment Result	Negative WordCloud	Positive WordCloud
Show	entries	Search:			
id	text				
1	king news				
2	mere days wisc parade massacre violent career criminal low cash bail				
3	apparently doesnt read brainwashes bird ted cruz fights qanon conspiracy theories jfk jr time dallas reason people				
4	liberals stupid simple math prisoners released ye				
5	job require gender studies degree				
6	im disappointed didnt tag aoc tweet				
7	bad socialism communism bird people hear ideas				
8	coordinated im pretty				
9	dear aoc attending friends hearing				
10	giveaway alert hotspawn teamed giveaway amazing gear bundle includes ultraw				

Gambar 3.6

Pada gambar 3.6 menampilkan bagian Cleaned Data, data disini sudah dibersihkan sehingga hanya menampilkan tweet saja tanpa informasi akun, mention atau link pada tweet.

All Data		Cleaned Data	Sentiment Result	Negative WordCloud	Positive WordCloud
Show	entries	Search:			
id	text	sentiment			
1	@DavidH87254937 @CNN @CityofQuincy @AOC You have the king for news https://t.co/IFibGWm2cv	Positive			
2	RT @MrAndyNgo: Mere days after the #Waukesha, Wisc. parade massacre, in which a violent career criminal was let out on low cash bail & went	Positive			
3	@yungb30599632 @macsam117 @MikeSington a Apparently doesn't read. Brainwashes himself on Big Bird / Ted Cruz fights, Q-Anon conspiracy theories, and probably thinks JFK Jr. is going to show up any time now in Dallas... You can't reason with these people.	Positive			
4	RT @The_FJC: Liberals are so stupid... They can't even do simple math... For example, they want to have all the prisoners released in 10 ye	Negative			
5	RT @PlanningMyPast: @KarenKnightSW @AOC What job could require a gender studies degree?	Positive			
6	@KarenKnightSW @ConceptualJames Im a little disappointed you didn't tag AOC in this tweet.	Negative			
7	RT @Stonekettle: How can we make this into a bad thing?! Socialism? Communism? Big Bird? What? COME ON, PEOPLE, LET'S HEAR SOME IDEAS! -- O	Negative			
8	It was coordinated. I'm pretty sure https://t.co/FkTZBQLBLx	Negative			
9	Dear AOC, Are you attending your friends hearing right now? https://t.co/DYojMJheNR	Positive			
10	RT @HotspawnEsports: Giveaway Alert s Hotspawn has teamed up with @AOC_Gaming to giveaway an amazing gear bundle, which includes an ultra w	Positive			

Gambar 3.7

Gambar 3.7 menampilkan bagian Sentiment Result yang memberikan informasi tweet tersebut negative atau positif.



Gambar 3.8

Pada gambar 3.8 menampilkan negative wordcloud dan positive wordcloud, terdapat diagram dan kata pada tweet yang sering atau yang sering digunakan pada akun akun twitter yang mentweet tentang Ted Cruz dan AOC

4. KESIMPULAN

Tweet seputar pemerintahan di Amerika Serikat meliputi Ted Cruz dan AOC memiliki banyak pandangan baik positive bahkan negative. Dari data yang dikumpulkan oleh praktikan menyimpulkan bahwa sebagian besar data terdiri atas opini positive yang diberikan oleh masyarakat seluruh dunia. Dengan program yang praktikan buat kali ini memudahkan user membaca data tanpa menggunakan waktu yang lama.

Link Youtube : https://youtu.be/_arVociZsV8

Link Github : <https://github.com/Qiraha/Project-DS>