

**LAPORAN PROYEK AKHIR  
PRAKTIKUM DATA SCIENCE**

**ANALISIS SENTIMEN MASYARAKAT  
TERHADAP PERUBAHAN IKLIM PADA MEDIA  
SOSIAL**



DAFFA NABIL LIBRIANA      123190085  
MUHAMMAD LATHIF H. A. M. 123190106

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

## **1. PENDAHULUAN**

Iklim adalah rata-rata cuaca dimana cuaca merupakan keadaan atmosfer pada suatu saat di waktu tertentu. Iklim didefinisikan sebagai ukuran rata-rata dan variabilitas kuantitas yang relevan dari variabel tertentu (seperti temperatur, curah hujan atau angin), pada periode waktu tertentu, yang merentang dari bulanan hingga tahunan atau jutaan tahun. Iklim berubah secara terus menerus karena interaksi antara komponen-komponennya dan faktor eksternal seperti erupsi vulkanik, variasi sinar matahari, dan faktor-faktor disebabkan oleh kegiatan manusia seperti misalnya perubahan penggunaan lahan dan penggunaan bahan bakar fosil. Perubahan Iklim adalah perubahan signifikan kepada iklim, suhu udara dan curah hujan mulai dari dasawarsa sampai jutaan tahun. Perubahan iklim terjadi karena meningkatnya konsentrasi gas karbon dioksida dan gas-gas lainnya di atmosfer yang menyebabkan efek gas rumah kaca.

Pada dasarnya, Gas Rumah Kaca dibutuhkan untuk menjaga suhu bumi tetap stabil. Akan tetapi, konsentrasi Gas Rumah kaca yang semakin meningkat membuat lapisan atmosfer semakin tebal. Penebalan lapisan atmosfer tersebut menyebabkan jumlah panas bumi yang terperangkap di atmosfer bumi semakin banyak, sehingga mengakibatkan peningkatan suhu bumi, yang disebut dengan pemanasan global. Perubahan iklim berdampak sangat luas pada kehidupan masyarakat. Kenaikan suhu bumi tidak hanya berdampak pada naiknya temperatur bumi tetapi juga mengubah sistem iklim yang mempengaruhi berbagai aspek pada perubahan alam dan kehidupan manusia. Beberapa contoh dampak negative perubahan iklim adalah gagal panen, cuaca ekstrim, dan meningkatnya wabah penyakit.

Dengan kemajuan teknologi, media sosial menjadi tempat bagi seseorang untuk menyampaikan ide maupun opini sehingga diketahui orang lain. Salah satu media sosial yang sering digunakan adalah Twitter. Banyak opini terkait perubahan iklim yang tersebar di Twitter dan peneliti ingin menganalisis sentimen opini tersebut apakah termasuk sentimen positif, negatif, maupun netral. Dengan diadakannya penelitian ini diharapkan dapat memberi informasi berguna bagi masyarakat maupun lembaga terkait untuk peningkatan pelestarian lingkungan.

## 2. METODE

### 2.1. *Business Understanding*

*Business Understanding* adalah tahap pertama yang membutuhkan pengetahuan objek bisnis, cara mendapatkan data, dan mencocokkan tujuan permodelan dengan tujuan bisnis agar model terbaik dapat diciptakan.

### 2.2. *Data Preparation*

*Data Preparation* atau bisa disebut juga dengan *data preprocessing* adalah suatu proses/langkahyang dilakukan untuk membuat data mentah menjadi data yang berkualitas

### 2.3. *NRC Sentiment Analysis*

*Sentiment analysis* adalah proses penggunaan *text analytics* untuk mendapatkan berbagai sumber data dari internet dan beragam platform media sosial. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada platform tersebut. Pada metode NRC, data diklasifikasikan menjadi 10 kategori (*positive, negative, anticipation, trust, joy, anger, fear, sadness, disgust, surprise*).

### 2.4. *Data Visualization*

Menurut Tableau, visualisasi data atau *data visualization* adalah tampilan berupa grafis atau visual dari informasi dan data. Dengan kata lain, *data visualization* mengubah kumpulan data menjadi hal lebih sederhana untuk ditampilkan.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Business Understanding

Pada kasus ini kami menggunakan sentimen masyarakat terhadap perubahan iklim pada media sosial twitter. Dataset ini didapatkan dari data yang dikumpulkan oleh Edward Qian dengan link <https://www.kaggle.com/edqian/twitter-climate-change-sentiment-dataset>. Dataset ini berisi 43.943 baris dan 3 kolom yaitu kolom *sentiment*, *message*, dan *tweetid*.

	sentiment	message	tweetid
1	-1	@tiniebeany climate change is an interesting hustle as it wa...	7.929274e+17
2	1	RT @NatGeoChannel: Watch #BeforeTheFlood right here, as...	7.931242e+17
3	1	Fabulous! Leonardo #DiCaprio's film on #climate change is ...	7.931244e+17
4	1	RT @Mick_Fanning: Just watched this amazing documentary...	7.931246e+17
5	2	RT @cnalive: Pranita Biswasi, a Lutheran from Odisha, gives ...	7.931252e+17
6	0	Unamshow awache kujinga na iko global warming https://t....	7.931254e+17
7	2	RT @cnalive: Pranita Biswasi, a Lutheran from Odisha, gives ...	7.931254e+17
8	2	RT @CCIRiviera: Presidential Candidate #DonaldTrump is Å€...	7.931266e+17
9	0	RT @AmericanIndian8: Leonardo DiCaprio's climate change ...	7.931271e+17
10	1	#BeforeTheFlood Watch #BeforeTheFlood right here, as @L...	7.931273e+17
11	1	RT @DrDeJarnett: It's vital that the public health community...	7.931279e+17
12	1	Bangladesh did not cause climate change, so the country do...	7.931297e+17
13	1	RT @NatGeoChannel: Watch #BeforeTheFlood right here, as...	7.931299e+17
14	1	Now's the time: we need a strong #FTT that works for those...	7.931300e+17
15	2	RT @WorldResources: Reflections on Leonardo DiCaprio's cl...	7.931302e+17
16	1	RT @littoralsociety: The problem with global warming's c...	7.931304e+17
17	1	RT @NatGeoChannel: One of the easiest ways to help comb...	7.931310e+17

#### 3.2. Data Preparation

##### 3.2.1. Cleaning Data

Proses *cleaning data* atau pembersihan data dilakukan dengan membersihkan kolom *message* dari entitas-entitas twitter yang tidak diperlukan dalam proses *sentiment analysis*. Entitas-entitas yang tidak dibutuhkan adalah emoji, *retweet*, *tagging*, nomor, tautan, dan spasi yang tidak penting.

	sentiment	message	tweetid
1	anti	climate change is an interesting hustle as it was global war...	7.929274e+17
2	pro	watch beforetheflood right here astravels the world to tackl...	7.931242e+17
3	pro	fabulous leonardo dicaprios film on climate change is brillia...	7.931244e+17
4	pro	just watched this amazing documentary by leonardodicapri...	7.931246e+17
5	news	pranita biswasi a lutheran from odisha gives testimony on ef...	7.931252e+17
6	neutral	unamshow awache kujinga na iko global warming	7.931254e+17
7	news	presidential candidate donaldtrump is dangerous on climat...	7.931266e+17
8	neutral	leonardo dicaprios climate change documentary is free for a...	7.931271e+17
9	pro	beforetheflood watch beforetheflood right here astravels th...	7.931273e+17
10	pro	its vital that the public health community addresses climate ...	7.931279e+17
11	pro	bangladesh did not cause climate change so the country do...	7.931297e+17
12	pro	nows the time we need a strong ftt that works for those har...	7.931300e+17
13	news	reflections on leonardo dicaprios new climate change film b...	7.931302e+17
14	pro	the problem with global warmingand the reason it continue...	7.931304e+17
15	pro	one of the easiest ways to help combat climate change in y...	7.931310e+17
16	pro	watch leo dicaprios climate change doc online for free befor...	7.931325e+17
17	pro	heres how climate change will affect sea level rise and flood...	7.931326e+17

### 3.2.2. Tokenisasi

Setelah *data cleaning* dilakukan, proses selanjutnya adalah tokenisasi. Pada tahap ini data dipisahkan menjadi per kata. Pemisahan kata ini bertujuan untuk proses *sentiment analysis* di tahap selanjutnya.

	sentiment	tweetid	word
1	anti	7.929274e+17	hustle
2	anti	7.929274e+17	global
3	anti	7.929274e+17	warming
4	anti	7.929274e+17	planet
5	anti	7.929274e+17	stopped
6	anti	7.929274e+17	warming
7	anti	7.929274e+17	foryes
8	anti	7.929274e+17	suv
9	anti	7.929274e+17	boom
10	pro	7.931242e+17	watch

### 3.3. NRC Sentiment Analysis

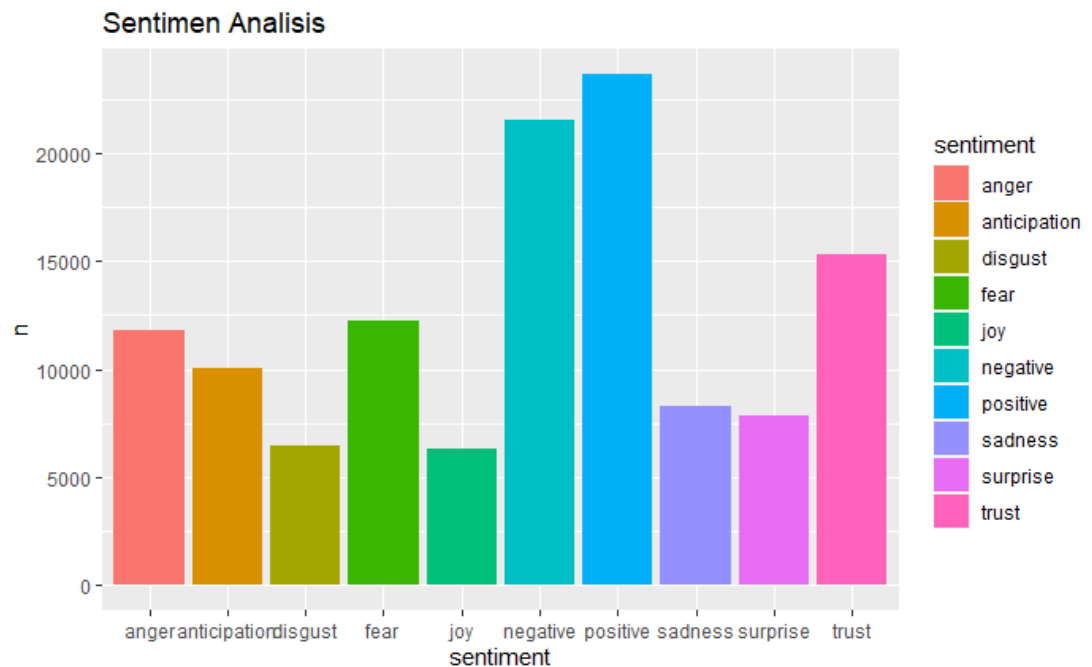
Setelah data dipisahkan pada proses tokenisasi, selanjutnya adalah analisis sentimen. Pada tahap ini tiap kata dikategorikan menggunakan metode NRC menjadi 10 kategori dan didapatkan hasil sebagai berikut.

	sentiment	n
1	positive	23672
2	negative	21594
3	trust	15358
4	fear	12214
5	anger	11791
6	anticipation	10071
7	sadness	8315
8	surprise	7845
9	disgust	6424
10	joy	6282

### 3.4. Data Visualization

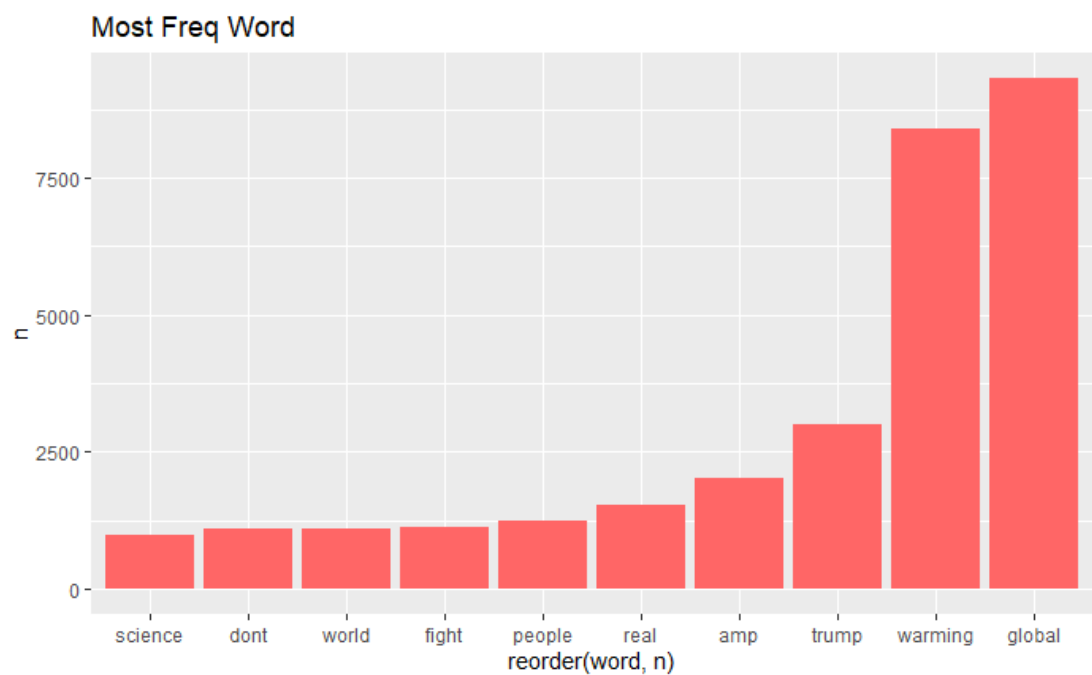
#### 3.4.1. Sentiment Plot

Sentiment plot adalah visualisasi data berdasarkan jumlah kata pada tiap kategori NRC.



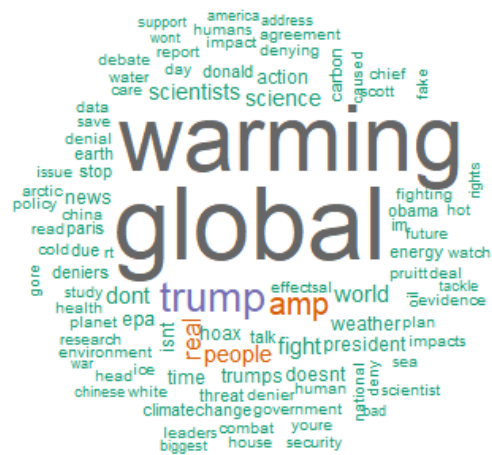
#### 3.4.2. Freq Word

Freq word adalah visualisasi data berdasarkan frekuensi kata yang muncul pada data.



### 3.4.3. Word Cloud

*Word cloud* (awan kata) atau sering disebut juga sebagai *tag cloud* merupakan representasi visual dari data teks. Frekuensi data yang paling besar akan memiliki ukuran yang paling besar pula.



#### 4. KESIMPULAN

Pada project ini dapat disimpulkan bahwa dari 43.943 data sentimen yang ada, diperoleh hasil *sentiment analysis* berupa 12.214 kata berkategori *fear*, 23.672 kata berkategori *positive*, 21.594 kata berkategori *negative*, 15.358 berkategori *trust*, 11.791 kata berkategori *anger*, 10.071 kata berkategori *anticipation*, 8.315 kata berkategori *sadness*, 7.845 kata berkategori *surprise*, 6.424 kata berkategori *disgust*, dan 6.282 kata berkategori *joy*. Kata dengan frekuensi terbanyak yaitu *global*, dan kategori kata dengan jumlah terbanyak yaitu *positive*.