

MILESTONE 1

oleh Kelompok 6 :

Muhamad Musta'in

Muzzammil Fadli

Nashrul Fatah

Novi Dwiasih

Nurul Hestinintyas

Topik : Climate Change

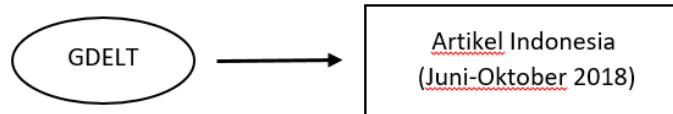
A. DESKRIPSI PROJECT

Bumi satu-satunya tempat tinggal bagi manusia kini telah banyak berubah. Perubahan itu terjadi dikarenakan banyak hal ada faktor alam dan faktor manusia. Salah satu yang terjadi adalah perubahan iklim. Iklim adalah rata-rata cuaca dimana cuaca merupakan keadaan atmosfer pada suatu saat di waktu tertentu. Iklim juga didefinisikan sebagai ukuran rata-rata dan variabilitas kuantitas yang relevan dari variabel tertentu (seperti temperatur, curah hujan atau angin), pada periode waktu tertentu. 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 fosil.

Perubahan iklim ini berpengaruh kepada banyak aspek seperti ketahanan pangan, ekonomi, kesehatan, sosial, dan politik. Dalam ketahanan pangan, perubahan iklim mempengaruhi hasil pertanian dan peternakan. Dalam kesehatan perubahan iklim yang terjadi dari titik ekstrem ke titik ekstrem lainnya berpengaruh pada tubuh manusia. Masih banyak lagi contoh kasus lain yang disebabkan oleh perubahan iklim. Semua yang terjadi tersimpan dalam data sehingga dapat kita analisis bagaimana dampak perubahan iklim terhadap kehidupan manusia

A. DATASET

Dataset yang didapatkan dari GDELT (Global Database of Events, Language, and Tone) berupa parameter/unsur yang terdapat dalam suatu artikel. Dalam proyek ini hanya difokuskan pada peristiwa yang terjadi di Indonesia, dengan memilih kode Negara (ID) pada periode waktu Juni-Oktober 2018.



Dari kumpulan artikel peristiwa di Indonesia pada periode tersebut dilakukan filter untuk memilih artikel yang membahas mengenai Perubahan Iklim (Climate Change). Beberapa parameter dari GDELT yang akan digunakan untuk analisis perubahan iklim diantaranya :

- | | |
|--------------------|--------------------------|
| a. SQLDATE, | g. ActionGeo_FullName |
| b. Actor1Code, | h. ActionGeo_CountryCode |
| c. Actor1Name | i. ActionGeo_Lat |
| d. Actor1Type1Code | j. ActionGeo_Long |
| e. QuadClass | k. SOURCEURL |
| f. GoldsteinScale | |

DATE	Actor1Code	Actor1Name	Actor1Type1Code	QuadClass	GoldsteinScale	ActionGeo_FullName	ActionGeo_CountryCode	ActionGeo_Lat	ActionGeo_Long	
70901	COP	PRISON	COP	4	-5.0	Jakarta, Jakarta Raya, Indonesia	ID	-6.17444	106.829	http://vn
30901	None	None	None	1	2.8	Tamansari, Jawa Tengah, Indonesia	ID	-6.9556	110.057	http://vn
30901	None	None	None	1	1.0	Jakarta, Jakarta Raya, Indonesia	ID	-6.17444	106.829	http://vn
30901	None	None	None	1	7.0	Jakarta, Jakarta Raya, Indonesia	ID	-6.17444	106.829	http://vn
30901	None	None	None	4	-5.0	Jakarta, Jakarta Raya, Indonesia	ID	-6.17444	106.829	http://vn
30901	None	None	None	1	2.8	Tamansari, Jawa Tengah, Indonesia	ID	-6.9556	110.057	http://vn
30901	COP	POLICE CHIEF	COP	1	1.9	Lokasari, Jawa Barat, Indonesia	ID	-7.0588	108.42	http://vn

Pengelompokan kata kunci dalam analisis perubahan iklim :

Iklim/Unsur	Penyebab	Proses	Dampak	Mitigasi
Atmosfer: - Oksigen - Karbon - Karbondioksida - Karbonmonoksida - Ozon - Sinar matahari - Cuaca - Suhu - Celsius - Curah Hujan - Awan - Uap - Global Warming	Bumi : - Motor - Pembakaran - Alam - Hutan - Erupsi - Gunung api - Polusi - Efek rumah kaca - Plastik	- Kondensasi - Peningkatan - Penurunan - Dingin - Panas	- Makhluk hidup - Hewan - Manusia - Tumbuhan - Lingkungan - Laut - Air - Banjir - Kekeringan - Kutub - Tenggelam - Pulau	- Energi alternative - Hemat - Penghijauan - Pencegahan - Forum

B. EKSPLORASI DATA

1. Untuk melakukan analisa pada project ini akan dilakukan analisa terhadap beberapa parameter diantaranya :

- Analisa pada QuadClass untuk menganalisa jenis peristiwa yang terjadi.
- Analisa pada Goldsteinscale untuk menangkap dampak potensial teoritis yang akan dimiliki oleh jenis peristiwa pada stabilitas suatu negara
- Analisa Actor untuk mengetahui siapa saja yang berperan dalam peristiwa
- Analisa lokasi (Latitude, Longitude) untuk menganalisa daerah mana yang membahas tentang peristiwa yang diperlukan

Dari contoh keempat analisa parameter tersebut kemudian digolongkan atau dibuat plotting berdasarkan topic yang berhubungan dengan perubahan iklim. Dan bisa dilakukan analisa jumlah pemberitaan topic tersebut.

C. PERKEMBANGAN PROJECT

Data yang telah didownload dari GDELT kemudian diekstrak dengan difilter hanya diambil data *event* yang terjadi di Indonesia. Data yang telah diekstrak dan difilter disimpan dalam bentuk TSV (tab separated value), yang kemudian dilakukan *pickling data* yang sekaligus menyatukan keseluruhan TSV tersebut dalam satu dataframe.

Dataframe tersebut kemudian ditambahkan dua kolom baru yaitu “title” dan “content” untuk proses *scraping* yang akan mengambil judul dan memasukkannya ke kolom “title”, serta isi berita ke kolom “content”. *Scraping* nantinya dilakukan mengacu ke alamat url yang ada di kolom “SOURCEURL”. Dataframe tersebut kemudian dikonversi menjadi SQL dalam database baru bernama “news” dengan nama tabel “gdelt”, serta mengacu ke header file yang telah didownload dari GDELT.

Berikut adalah tangkapan layar dari tabel di HeidiSQL :

GLOBALEVENTID	SQLDATE	MonthYear	Year	FractionDate	Actor 1Code	Actor 1Name	Actor 1CountryCode	Actor 1KnownGroupCode	Actor 1EthnicCode
797,260,234	20181025	201810	2018	2018.8082	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
797,260,776	20181025	201810	2018	2018.8082	BUS	COMPANIES	(NULL)	(NULL)	(NULL)
797,260,783	20181025	201810	2018	2018.8082	BUS	COMPANIES	(NULL)	(NULL)	(NULL)
797,260,856	20181025	201810	2018	2018.8082	BUS	COMPANIES	(NULL)	(NULL)	(NULL)
797,261,026	20181025	201810	2018	2018.8082	CHN	CHINA	CHN	(NULL)	(NULL)
797,261,340	20181025	201810	2018	2018.8082	CVL	COMMUNITY	(NULL)	(NULL)	(NULL)
797,261,451	20181025	201810	2018	2018.8082	CVL	ENGINEER	(NULL)	(NULL)	(NULL)
797,261,906	20181025	201810	2018	2018.8082	GOV	MINIST	(NULL)	(NULL)	(NULL)
797,262,133	20181025	201810	2018	2018.8082	GOV	GOVERNMENT	(NULL)	(NULL)	(NULL)
797,262,309	20181025	201810	2018	2018.8082	IDN	INDONESIA	IDN	(NULL)	(NULL)
797,262,310	20181025	201810	2018	2018.8082	IDN	INDONESIA	IDN	(NULL)	(NULL)
797,262,311	20181025	201810	2018	2018.8082	IDN	INDONESIA	IDN	(NULL)	(NULL)
797,262,312	20181025	201810	2018	2018.8082	IDNGOVMED	INDONESIA	IDN	(NULL)	(NULL)
797,262,823	20181025	201810	2018	2018.8082	MED	SPOKESMAN	(NULL)	(NULL)	(NULL)
797,263,194	20181025	201810	2018	2018.8082	REL	ABBOT	(NULL)	(NULL)	(NULL)
797,265,247	20181025	201810	2018	2018.8082	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
797,265,248	20181025	201810	2018	2018.8082	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
797,265,919	20181025	201810	2018	2018.8082	COP	POLICE	(NULL)	(NULL)	(NULL)
797,266,314	20181025	201810	2018	2018.8082	FRA	FRENCH	FRA	(NULL)	(NULL)
797,266,753	20181025	201810	2018	2018.8082	IDN	INDONESIA	IDN	(NULL)	(NULL)
797,266,754	20181025	201810	2018	2018.8082	IDN	INDONESIA	IDN	(NULL)	(NULL)
797,266,755	20181025	201810	2018	2018.8082	IDN	EAST JAVA	IDN	(NULL)	(NULL)
797,266,756	20181025	201810	2018	2018.8082	IDN	INDONESIA	IDN	(NULL)	(NULL)
797,266,757	20181025	201810	2018	2018.8082	IDN	INDONESIA	IDN	(NULL)	(NULL)
797,266,758	20181025	201810	2018	2018.8082	IDN	INDONESIA	IDN	(NULL)	(NULL)

Scraping dilakukan untuk mengambil data yang lebih mendetail dari data yang telah disediakan di GDELT. Dengan membagi rentang waktu ke dalam beberapa bagian waktu ke beberapa anggota kelompok, harapannya data yang diproses tak terlampaui besar. Ke depannya, kami merencanakan untuk menganalisis lebih jauh melalui judul dan isi berita yang kami peroleh. Analisis lebih jauh dari data ini akan kami kerjakan dalam milestone berikutnya.

Berikut adalah contoh tangkapan layar proses *scraping* di Jupyter Notebook:

```
SELECT * FROM `gdelt` WHERE (TITLE IS NULL OR TRIM(TITLE) = '')
https://www.thechronicle.com.au/news/andrew-bolt-dump-him-scomo-hes-no-good-for-you/3558125/
https://www.businesstimes.com.sg/opinion/dont-let-the-weather-disrupt-your-business
http://www.theedgemarkets.com/article/inside-governmentrun-war-room-fighting-indonesian-fake-news
http://www.theedgemarkets.com/article/inside-governmentrun-war-room-fighting-indonesian-fake-news
http://www.mining.com/copper-price-drop-prompts-expansion-rethink-top-producer/
https://www.benarnews.org/english/news/indonesian/indonesia-politics-10242018161548.html
http://www.theedgemarkets.com/article/inside-governmentrun-war-room-fighting-indonesian-fake-news
http://www.theedgemarkets.com/article/inside-governmentrun-war-room-fighting-indonesian-fake-news
https://www.thechronicle.com.au/news/andrew-bolt-dump-him-scomo-hes-no-good-for-you/3558125/
http://www.theedgemarkets.com/article/inside-governmentrun-war-room-fighting-indonesian-fake-news
http://www.theedgemarkets.com/article/inside-governmentrun-war-room-fighting-indonesian-fake-news
https://www.thechronicle.com.au/news/andrew-bolt-dump-him-scomo-hes-no-good-for-you/3558125/
http://www.theedgemarkets.com/article/inside-governmentrun-war-room-fighting-indonesian-fake-news
https://www.benarnews.org/english/news/indonesian/indonesia-politics-10242018161548.html
https://www.thechronicle.com.au/news/andrew-bolt-dump-him-scomo-hes-no-good-for-you/3558125/
http://en.tempo.co/read/news/2018/10/25/057922846/Two-Malaysian-Drug-Smugglers-Arrested-in-Surabaya
http://en.tempo.co/read/news/2018/10/25/057922846/Two-Malaysian-Drug-Smugglers-Arrested-in-Surabaya
http://en.tempo.co/read/news/2018/10/25/057922846/Two-Malaysian-Drug-Smugglers-Arrested-in-Surabaya
```

Analisis paling sederhana yang bisa kami lakukan sejauh ini adalah menghitung berapa banyak berita mengenai *climate change* dalam tabel kami, menggunakan *keyword* yang kami definisikan di tabel pengelompokan kata kunci di subbab sebelumnya. Sebelumnya, kami juga mengekstrak *row* tertentu yang memuat kata kunci yang kami inginkan. Berikut adalah hasil tangkapan layarnya :

```
1 # ganti keyword di sini
2 keyword = "passengers", "climate"
3
4 if isinstance(keyword, str):
5     query = "SELECT * FROM gdelt WHERE content LIKE '%s'" % keyword
6 else:
7     query = "SELECT * FROM gdelt WHERE content LIKE '%s'" % (keyword[0])+" OR content LIKE '%s'"%(len(keyword)-1)
8     % (keyword[1:])
9 run(query)
```

untryCode	ActionGeo_ADM1Code	ActionGeo_Lat	ActionGeo_Long	ActionGeo_FeatureID	DATEADDED	SOURCEURL	title	content
ID	ID08	-7.1046	112.337	10205777	20181025	https://www.thechronicle.com.au/news/andrew-bo...	Don't be weak, just dump Turnbull	How does it help the Morrison Government to no...
ID	ID	-5	120	ID	20181025	https://www.businesstimes.com.sg/opinion/dont-...	Don't let the weather disrupt your business	WHEN it comes to the Asia-Pacific, it is not h...
ID	ID08	-7.1046	112.337	10205777	20181025	https://www.thechronicle.com.au/news/andrew-bo...	Don't be weak, just dump Turnbull	How does it help the Morrison Government to no...

Sebetulnya, dalam proses penanganan data ini, masih ada beberapa kendala. Pertama, kami masih belum tahu bagaimana caranya untuk menangani data yang *duplicate*. Beberapa *row* terlihat *duplicate* jika dibandingkan dari kolom *title* dan *content*nya, meskipun sebetulnya di kolom lain memiliki nilai yang berbeda. Sebetulnya, hal ini dapat diantisipasi hanya dengan memerhatikan kolom *title* dan *content* dan meniadakan *duplicate row*. Kedua, kami tidak memerhatikan apakah berita yang ikut terekstrak akan relevan sesuai *keyword* yang kami gunakan (khususnya akan membahas *climate change*).

Selanjutnya, kami juga menghitung berapa banyak berita mengenai *climate change* tersebut dengan mengelompokkannya berdasarkan tanggal *event* dari kolom “SQLDATE”. Berikut adalah hasil tangkapan layar sebagai contoh sederhana prosesnya :

Menghitung row yang hanya memuat keyword yang kita inginkan.

```
1 if isinstance(keyword, str):
2     query = "SELECT sqldate AS 'Tanggal Berita', count(*) AS 'Jumlah Berita' FROM gdelt WHERE content LIKE '%s%'"
3     %keyword + " GROUP BY sqldate"
4 else:
5     query = "SELECT sqldate AS 'Tanggal Berita', count(*) AS 'Jumlah Berita' FROM gdelt WHERE content LIKE '%s%'"
6     % (keyword[0]) + " OR content LIKE '%s%' " * (len(keyword)-1) % (keyword[1:]) + " GROUP BY sqldate"
7 run(query)
```

	Tanggal Berita	Jumlah Berita
0	20171026	4
1	20171028	2
2	20181025	93
3	20181026	49
4	20181027	58
5	20181028	14
6	20181029	1620

Dari dataframe baru tersebut, langkah selanjutnya yang kami lakukan adalah mengumpulkan seluruh dataframe yang kami kerjakan masing-masing untuk memperoleh satu dataframe berisi jumlah berita di kurun waktu Juni 2018 sampai Oktober 2018 yang telah terkelompokkan sesuai tanggal beritanya. Selanjutnya, dari dataframe tersebut dapat digunakan untuk membuat plot *time series* untuk melihat jumlah berita yang membahas *climate change* per tanggal dalam kurun waktu tersebut. Dengan demikian, kami dapat melihat *insight* secara umum apakah trennya positif (yang berarti makin sering diberitakan, dan disimpulkan menjadi permasalahan yang lebih pelik), atau biasa saja, atau negatif.