



Analisis Sentimen Twitter

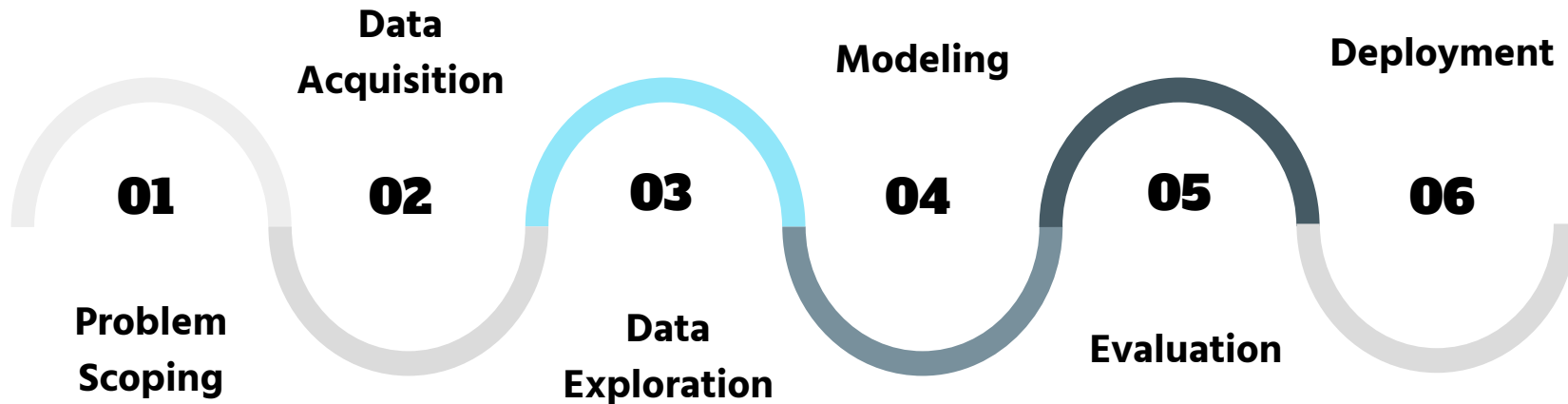
Terhadap “KTT Ke-17 G20 Indonesia”
Menggunakan Machine Learning

By Ahmad Kamiludin





Workflow



Problem Scoping

G20 adalah forum Internasional yang fokus pada koordinasi kebijakan di bidang ekonomi & pembangunan. Dimana G20 ini lahir sebagai respon atas krisis ekonomi dunia pada tahun 1997-1998. Dan untuk pertama kalinya, Indonesia memegang Presidensi G20 yang diselenggarakan di Bali pada tanggal 15-16 November 2022 kemarin.



Problem Scoping

Analisis Sentimen ini bisa digunakan untuk mengetahui dan memahami opini dan reaksi masyarakat terhadap **"G20 Indonesia"** melalui platform sosial media Twitter.



Data Acquisition

- Dataset yang di ambil adalah berupa Tweet Netizen **dalam negeri (Indonesia)** maupun Netizen **luar negeri**
- Data Tweet yang digunakan adalah yang **berbahasa Inggris** dan diambil menggunakan Bahasa Pemrograman Python
- Data Tweet di ambil dalam kurun waktu **1 tahun** yakni mulai tanggal **1 November 2021 – 1 November 2022**, dengan jumlah total data sebanyak **5000 Tweet**
- Data Tweet dianalisa menggunakan Software **KNIME** dan sedikit tambahan Bahasa Pemrograman Python



Data Acquisition

```
# IMPORT MODULES
import snsrape.modules.twitter as sntwitter
import pandas as pd
```

✓ 4.1s

```
# MENYIAPKAN SEGALA SESUATU YANG DIPERLUKAN : KATA KUNCI, BAHASA, JUMLAH TWEET, DLL.
query = "(g20indonesia) lang:en until:2022-11-01 since:2021-11-01"
tweets = []
limit = 5000
```

✓ 0.3s

```
# PROSES CRAWLING/NARIK DATA DARI TWITTER
try:
    print("Starting to scrape tweets")
    for tweet in sntwitter.TwitterSearchScrapper(query=query).get_items():
        if len(tweets) == limit:
            break
        else:
            tweets.append([tweet.date, tweet.user.username, tweet.content])
    df = pd.DataFrame(tweets, columns=['Datetime', 'Username', 'Tweet'])
except Exception as e:
    print(e)

print("finished scraping tweets")
```

Data Acquisition

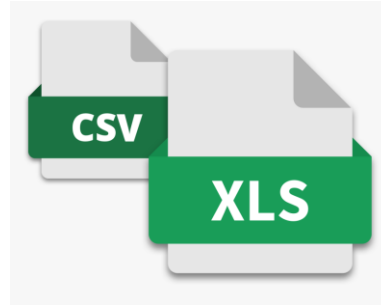
```
# SIMPAN KE CSV FILE
df = df.to_csv('g20indonesia_tweets.csv', index=False)
```

```
# MENAMPILKAN DATASET
read_df = pd.read_csv('g20indonesia_tweets.csv')
read_df
```

	Datetime	Username	Tweet
0	2022-10-31 22:44:10+00:00	m3trainer	🔥 @g20org @Cop27P We have had enough of witnes...
1	2022-10-31 22:31:41+00:00	anthony82149035	@tni blood thirty savages in uniform racist an...
2	2022-10-31 22:16:51+00:00	corsicanmafioso	🔊 @G7 @EUCouncil @g20org: The devastating floo...
3	2022-10-31 22:06:55+00:00	Ekklesia_co_uk	G20 Summit looks to world religions for soluti...
4	2022-10-31 21:48:37+00:00	prof_ahmetkuru	Indonesia is hosting G20 & its religious f...
...
4995	2022-09-29 14:48:59+00:00	herlan_sydney	.@Menlu_RI Civic space is crucial for sustaina...
4996	2022-09-29 14:40:28+00:00	herlan_sydney	🔊 @G7 @EUCouncil @g20org: The devastating floo...
4997	2022-09-29 14:38:29+00:00	TheVineetMittal	I congratulate Shri @amitabhk87 and @NITIAayog...
4998	2022-09-29 14:36:03+00:00	Kim83897697	🔥 @g20org @EUCouncil @Cop27P @vonderleyen @tru...
4999	2022-09-29 14:29:13+00:00	Kumar27147450	@IndianEmbJkt @g20org @MEAIndia @cgibali @Agri...

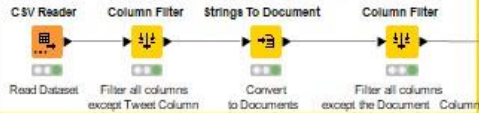
5000 rows × 3 columns

Data Acquisition



Workflow Sentiment Analysis Using KNIME

1.) Importing & Reading Dataset



2.) Text Cleaning



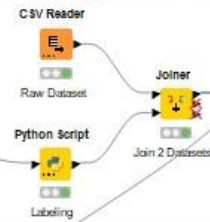
3.) Text Preprocessing 1



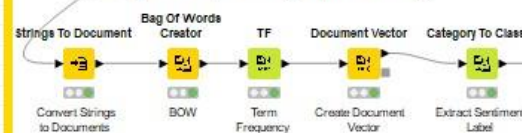
4.) Visualizations Using Word Cloud



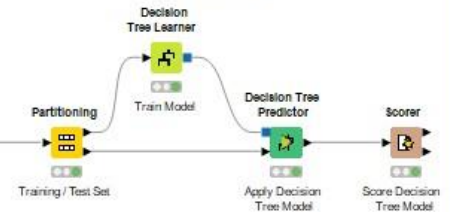
5.) Labelling Using Vader Sentiment



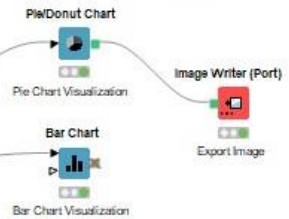
6.) Text Preprocessing 2 (For Modeling)



7.) Modeling



8.) Visualizations & Deployment



Data Exploration

- **Text Preprocessing** : Mempersiapkan teks yang **tidak terstruktur** menjadi data yang **baik dan siap untuk diolah**.
- Text Preprocessing yang digunakan antara lain :
 - Membersihkan teks dari tanda Hashtag, Emoji, URL, Special Characters, dan lain-lain.
 - Tokenisasi (setiap kata akan dipisah-pisah terlebih dahulu)
 - Memberi tanda pada suatu kata yang sudah dipisah-pisah menggunakan Part Of Speech (POS)
 - Menghilangkan tanda baca, kata penghubung, dan mengubah suatu kata ke kata dasarnya.
 - Mengubah data teks menjadi vektor sehingga bisa dipahami oleh komputer.
 - Menghitung frekuensi kemunculan kata pada seluruh dokumen.

Data Exploration



- **Labeling** : Proses pelabelan atau memberi label pada setiap Tweet dan menjadi 3 label, yaitu **Positif, Netral, dan Negatif**.
- Proses pelabelan menggunakan Metode **VADER**, yang merupakan salah satu metode Labeling untuk data teks pada Social Media.
- Metode **VADER** memanfaatkan **Skor sentimen** pada suatu teks, yang **diperoleh** dengan **menjumlahkan intensitas setiap kata** dalam teks tersebut.
- Misalnya, kata-kata seperti **"LOVE", "HAPPY", "WONDERFUL", "ENJOY"**, semuanya termasuk sentimen **Positif**. Sebaliknya, kata-kata seperti **"BAD", "CORRUPT", "DAMAGE"** dianggap sebagai pernyataan **Negatif**.





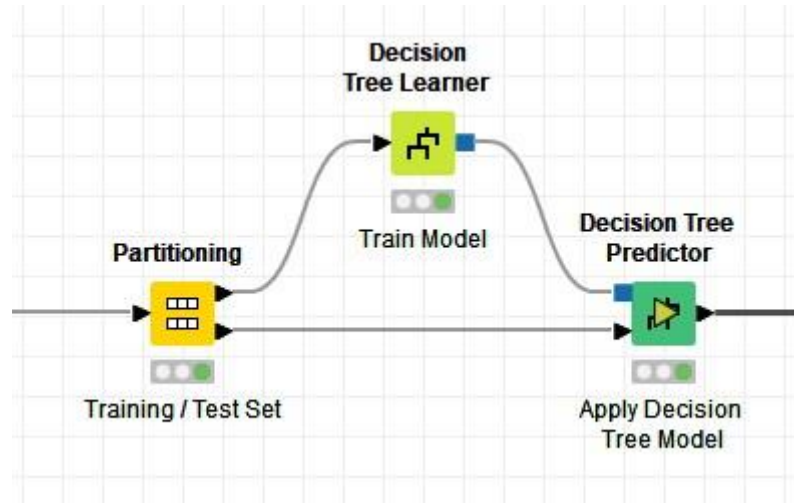


Berdasarkan hasil visualisasi diatas, selain banyak membicarakan **Indonesia** selaku Presidensi G20 tahun 2022, orang-orang juga banyak membicarakan tentang **Rakyat, Krisis, Iklim, Fasilitas, Kelaparan, Kerusakan, Kehidupan,** dan lain sebagainya yang tentunya sangat erat kaitannya dengan G20 Indonesia.

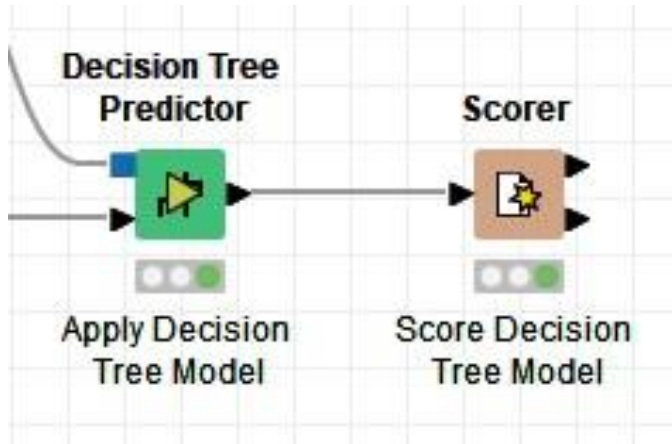


Modeling

Algoritma Machine Learning yang digunakan adalah **“Decision Tree”**, yang digunakan untuk **mengklasifikasikan Tweet**, mana yang termasuk **Positif**, **Netral**, maupun **Negatif**. Nantinya, kita bisa **memanfaatkan Model ini**. Sehingga, jika ada Tweet baru, **tidak perlu** lagi dilakukan klasifikasi Tweet secara **manual**.



Evaluation



Confusion Matrix - 3:59 - Scorer (Score Decision)

File Hilite

Document class \ Prediction (Document class)	Neutral	Positive	Negative
Neutral	142	33	4
Positive	52	324	11
Negative	36	21	377

Correct classified: 843 Wrong classified: 157

Accuracy: 84.3% Error: 15.7%

Cohen's kappa (κ): 0.756%

Berdasarkan hasil evaluasi, nilai Accuracy sebesar **84,3%**. Maka dari itu, Model ini sudah **cukup baik** dalam melakukan klasifikasi Tweet.

Evaluation

Username	Tweet	Vader Sentiment	Prediction
EINWarming	ANALYSIS-Indonesia to strike green energy deal with rich nations at G20 summit https://t.co/oEsyTZ8Ob8	Positive	Positive
KaylaDworin5	Aceh Support G20 #G20Indonesia #RecoverTogetherRecoverStronger #G20Indonesia2022 #GroupOfTwentyIndonesia https://t.co/0wfHvD0Rv	Positive	Positive
thebali_sun	Over 100 Additional Staff Added To Bali Airport Immigration Ahead Of G20 Summit https://t.co/N094TUD6lR #bali #airport #G20Indonesia	Neutral	Neutral
g20org	AMR affects all countries and resistance can spread across national borders and boundaries, collectively earning resistant infections the moniker "the silent pandemic". #G20Indonesia https://t.co/GClm8g3Eg9	Neutral	Neutral

Evaluation

Username	Tweet	Vader Sentiment	Prediction
sigmundsdroid	@SenatorWong How can Australia attend the @g20org with an active war criminal.	Negative	Negative
MaxStreit2	@EMM_1969 @KyivIndependent @G7 @Europarl_EN @g20org @UN @ForeignChinese @narendramodi @AP @BBCWorld I believe Ukraine is trading with Saudi Arabia who are currently conducting military operations in Yemen killing thousands of civilians and commuting warcrimes. Is Ukraine in the wrong side of history now?	Negative	Negative
euroscot1	@EdwardGLuce @FinancialTimes Should G20 approve the US's political bets on regime change in China & Russia? Indonesia meeting will be busy: the group's unsustainable governance of global finance post 2008. Voting: the WH may 'friendshore' some autocrats; China pressuring Indonesia gathering too. WCOA?	Negative	Positive

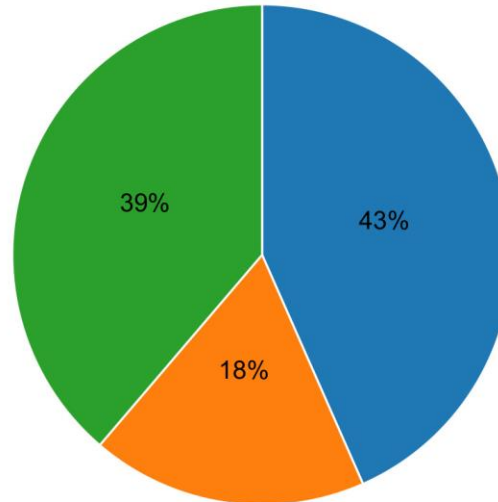
Deployment

Berikut Hasil Analisis Sentimen Terhadap **"KTT G20 Indonesia"**
Menggunakan Machine Learning:

Sentiment Analysis

G20 INDONESIA

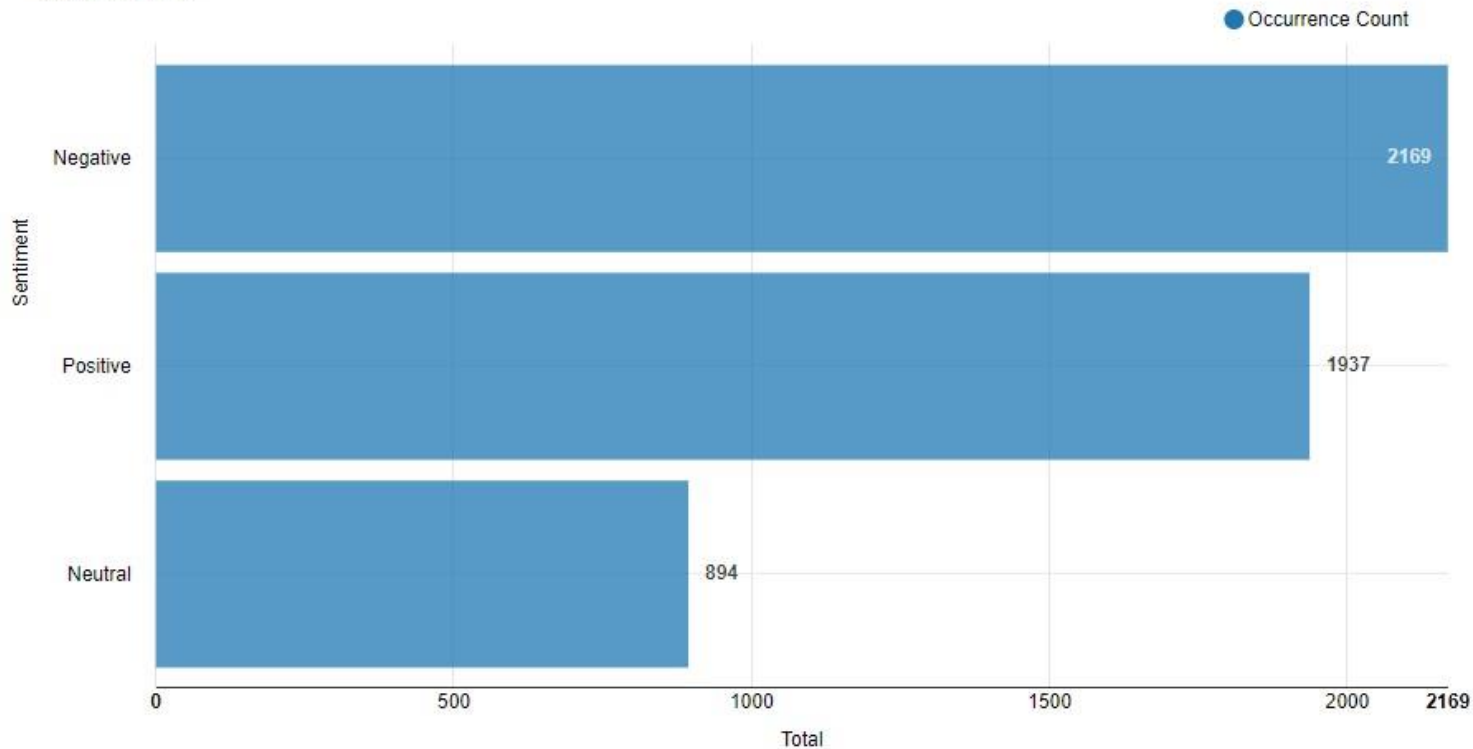
● Negative ● Neutral ● Positive



Deployment

Sentiment Analysis

G20 INDONESIA





Thank You!

Ahmad Kamiludin
Balung, Jember, Jawa Timur
My Others Portfolio :
<https://kamiludin-portfolio.my.id/>