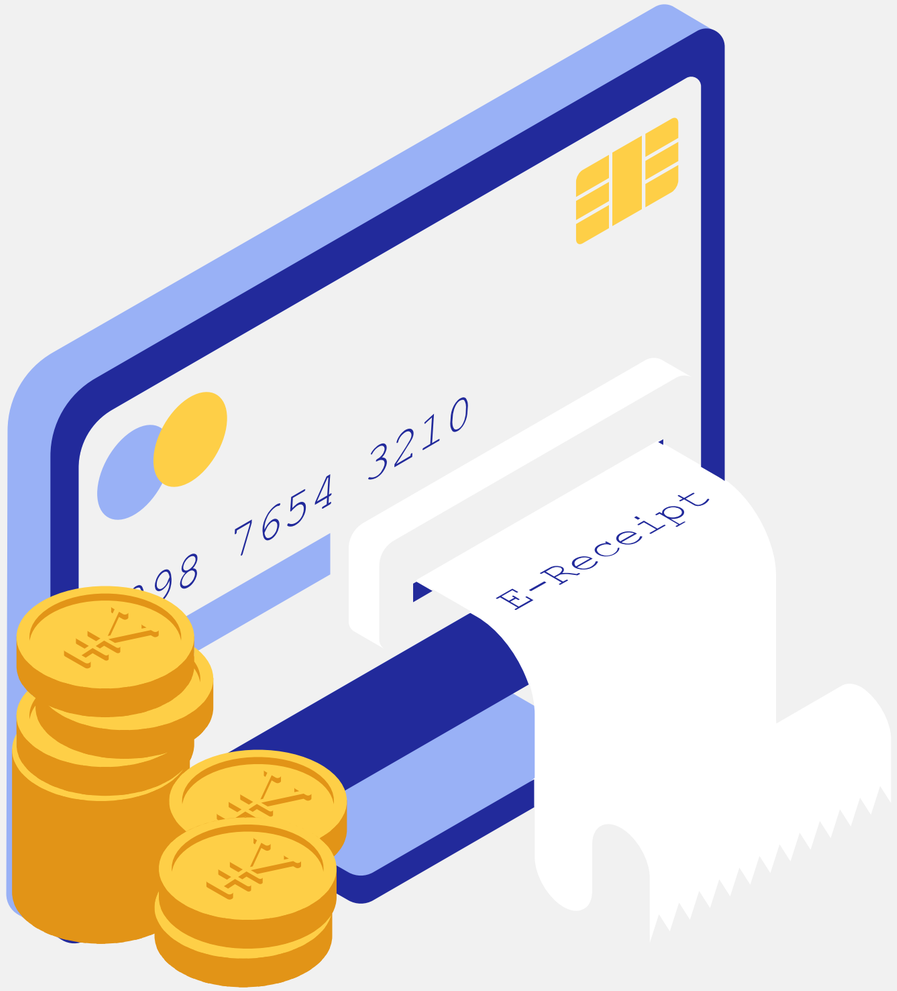
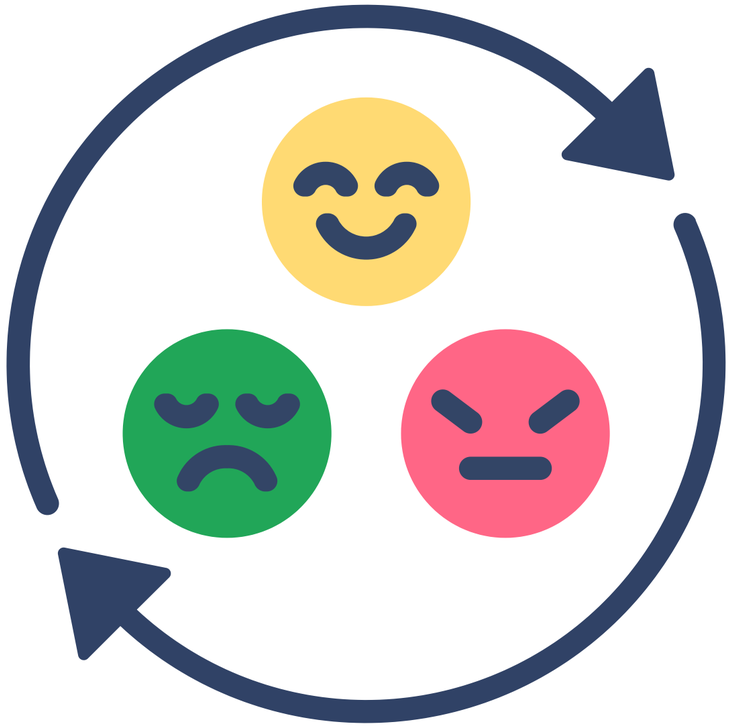


Understanding its importance and influence

Bank Mega Social Media Sentiment Analysis (SMSA)

How it changes the way client view our bank



Apa itu Sentimen Analisis?

Knowing how someone feels through text.

Analisis Sentimen adalah proses mendeteksi reaksi pelanggan terhadap suatu produk, merek, situasi, atau peristiwa melalui teks, postingan, ulasan, dan konten digital lainnya. Dengan menggunakan analisis sentimen, bisnis dapat memperoleh wawasan mendalam tentang bagaimana pelanggan mereka berpikir dan merasakan.

Mengapa Social Media?

- Jumlah konten yang luas
- Data Real-Time
- Demografi yang luas
- Ekspresi yang autentik

Mengapa Sentimen Analisis Penting?

Gaining deep insight into how customers think and feel.

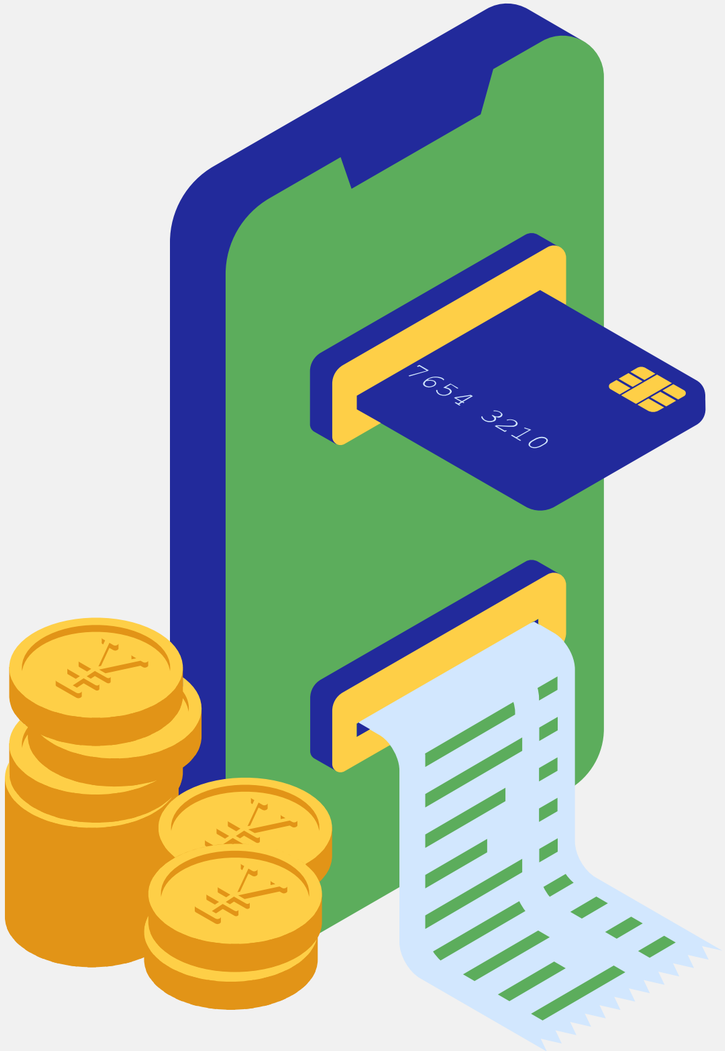
Perbankan adalah salah satu industri dalam sektor jasa di mana bisnisnya sangat bergantung pada bagaimana nasabah atau pelanggan berespon dan menilai pelayanan yang diberikan.

Mengenali dan memahami perasaan nasabah dapat menjadi faktor kunci dalam mencapai tujuan bisnis seperti menjaga loyalitas, meningkatkan kualitas pelayanan, dan memperbaiki pengalaman nasabah.

What will I do?

Social Media Scraping

Mengambil dan mengumpulkan informasi yang tersedia secara publik dari situs web media sosial, seperti kiriman, komentar, suka, pengikut, dan profil pengguna secara otomatis.



Sentiment Analysis Model Enhancement

Meningkatkan kinerja, akurasi, dan ketahanan model analisis sentimen yang sudah ada dengan mencoba berbagai teknik dan strategi untuk meningkatkan kemampuan model dalam mengidentifikasi dan mengklasifikasikan sentimen secara akurat pada data media sosial.

Project Milestone

Riset	EDA	Data preprocessing	Training model
1 Minggu	1 Minggu	3 Minggu • Data Labeling • Data Cleaning • Data Encoding	1 Minggu

Evaluasi

2 Minggu

Menganalisa performa dan akurasi model serta mencari opsi solusi yang dapat diterapkan untuk meningkatkan akurasi.

Inferensi & Fine-tuning

1 Minggu

Proses dapat berulang sesuai hasil evaluasi.

Purposed Model

Naive Bayes

Long Short Term Memory (LSTM)

Gated Recurrent Unit (GRU)

Fine-tune IndoBERT

Fine-tune RoBERTa

Fine-tune GPT 2

Laporan Magang

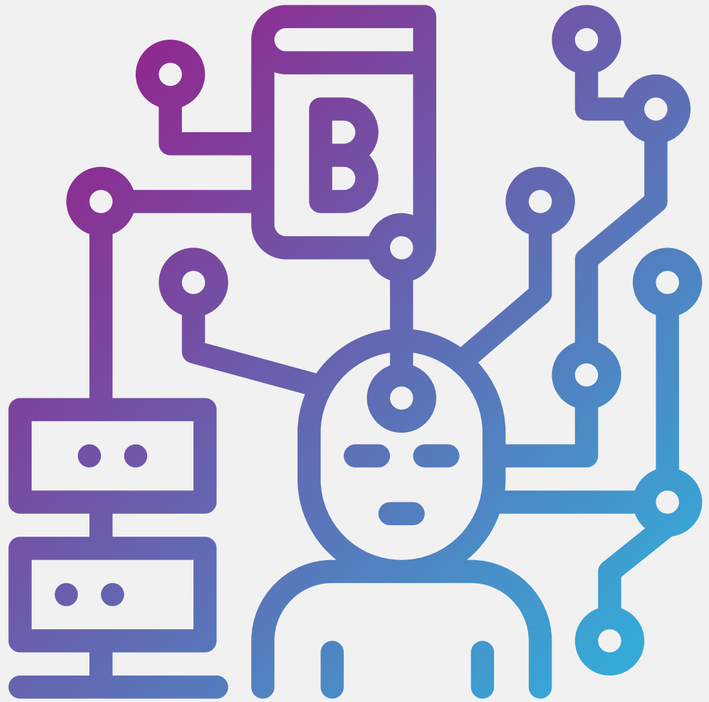
Social Media Scraping

```python
class InstagramScraper:
    def __init__(self, driver, today=datetime.date.today()):
        self.driver = driver
        self.cutoff_awal = None
        self.cutoff_akhir = None
        self.comment_awal = None
        self.comment_akhir = None
        self.today = today
        self.get_cutoff_comment_time(self.today)
```

```python
class FBScraper:
    def __init__(self, driver):
        self.driver = driver
        self.df_url = []
        self.url = []

        self.user_names=[]
        self.user_comments=[]
```

```python
class TwitterScraper:
    def __init__(self):
        self.result_links = set()
        self.result = set()
        self.set_tweets = set() #link tweet acc
        # PROXY = "116.254.117.162"
        self.option = Options()
```

Sentiment Analysis Modeling



Data yang Digunakan

1. df_sentiment_modified

	A	B	C	D	E
1	user	text	sentiment	**sentiment_mod**	
2	Reivolusi	Kenapa sulit banget Nelpon Call Center kartu kredit Bank Mega, #BankMega @BankMegaID	-1	0	
3	muh_andiafif	@BankMegaID yang terhormat ibu aerin di bankmega Jakarta, cara ibu menagih utang kartu kredit keponakan sy, ke istri sy, dg telpon berkali2 ke tempat kerjanya, telah membuat a'ib bagi istri sy. SANGAT KETERLALUAN!	-1	0	
4	Promo_Banter	@BankMegaID Silahkan tanya yg brsngktn yg mengatasnamakan BankMega @OJKRI @bankindonesia @jokowi semua kontak Teman Medsos org tsbt dihubungi, niat mempermalukan? @ https://t.co/QEOUXzXdFZ	-1	0	
5	radar_malang	Eks Bank Mega Malang Terancam Hadapi Tuntutan Ganda @infomalang #bankmega https://t.co/3ogW1knYF1	-1	0	
6	BacaDiBaBe	Deposito Rp 65 M di Bank Mega Raib, Nasabah Cerita Ada Kejanggalan pada 2012 #Deposito #BankMega #Sukubunga #tabungan #Bareskrim https://t.co/4zp7BPMyJH https://t.co/PBxaiSWapR	-1	0	
7	BacaDiBaBe	Dana Nasabah Rp 56 M di Bank Mega Raib, OJK: Pelanggar Akan Kena Sanksi #Jasakeuangan #BankMega #Bareskrim #OJK https://t.co/7X0seZBvXV https://t.co/TxqMjxWxFo	-1	0	
8	trenasia_com	Raib, Deposito Nasabah Bank Mega Syariah Rp20 Miliar! @megasyariah1 #kronologi #kejadian #bankmega #bankmegasyariah https://t.co/Gz7wd86cmT	-1	0	
9	nusabalicom	Kacab Bank Mega Bobol Deposito Nasabah Rp 62M https://t.co/CJwtPiE2Pt #berita #nusabali #KejariDenpasar #BankMega #Penggelapan #Penyelewengan	-1	0	
10	hendyk_eko	@bankmega dalang tuek turkmenistan	-1	0	
11	hendyk_eko	@bankmega dekeng bayar ONE ENVIROFMENT	-1	0	
12	DmenkBudi	Saya cukup perihatin dgn sikap dan gaya penagihan dept collector sekarang, perusahaan sebesar bank Mega @bankmega memperkerjakan orang2 dengan etika yg sangat buruk berbicara sangat tidak sopan.@ojk_ @bank_indonesia	-1	0	
13	Michii8143169	Sebagian recent log dri debt coll @BankMegaID yg ngutang siapa yg di cari siapa , DCnya gak sopan lagi #bankmega #debtcoll #rude #gaksopan #mengganggu https://t.co/NqLu4JjBmf	-1	0	
14	Michii8143169	Oiiii @BankMegaID kesabaran udah abs nihhh ! Debt coll nya kaya 💩 💩 💩 💩 💩 bgt dha ! Di Dm bilang mau di tindak lanjutin .. eh kenyataannya kaga ada kelanjutan #bankmega #rude #debtcoll #kurangajar #uneducated #bankindonesia #ojk #penagihutang #gakberadab #transmart #transcorp	-1	0	
15	rizchaniago	@BankMegaID @yunitanovalia Div IT nya @BankMegaID emg bloon.. masa subscribe news letter bisa klik/online, tp unsubscribenya harus lewat nelp. Pemaksaan yg terselubung. Gampangnya, blok aja semua email bankmega yg masuk @yunitanovalia ...	-1	0	
16	M19Tris	Penutupan kartu kredit bank Mega @BankMegaID #bankmega #kartukredit @ojkindonesia https://t.co/UmnPYy320V	0	1	

mega | btpn | btn

Data yang Digunakan

2. df_all

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		created	id	usernar	locatic	ends_c	lowers_	text	retweet_coun	favorite_cou	place	catego	sentiment_ori	sentiment_mod
2	0	2021-04-1	1,4E+18	jesicaprilli	East Java,	577	667	@JeniusConnect	0	0		bank btpn	1	2
3	1	2021-09-2	1,4E+18	JeniusConnect		78	82251	@dheeamalia Ba	0	0		bank btpn	99	99
4	3	2021-12-0	1,5E+18	JeniusConnect		78	82251	@Tiacid Hi, Tia. F	0	1		bank btpn	99	99
5	4	2021-10-1	1,4E+18	muwiosigu	9xliner	96	89	@CaratShopINA	0	0		bank btpn	99	99
6	5	2021-05-0	1,4E+18	hahihahiha	Indonesia	1000	645	@JeniusConnect	0	0		bank btpn	1	2
7	7	2021-04-1	1,4E+18	Ariyadi	Yogyakart	1255	524	@jeniushelp Sud	0	0	Place(_api	bank btpn	-1	0
8	8	2021-06-2	1,4E+18	JeniusConnect		78	82251	@tri_moel Sebag	0	0		bank btpn	99	99
9	9	2021-06-3	1,4E+18	adm_shim	Jakarta Pu	17	1588	@dagangkorea É	0	0		bank btpn	99	99
10	11	2021-07-1	1,4E+18	Lisaaamor	Indonesia	226	211	@theresiaavila @	0	0		bank btpn	99	99
11	12	2021-11-1	1,5E+18	menerjang	Jakarta Ca	508	874	Oh gue ternyata	1	1		bank btpn	99	99
12	13	2021-07-2	1,4E+18	princessra	Klinik cinta	670	1152	Dear @JeniusCo	0	0		bank btpn	-1	0
13	14	2021-09-0	1,4E+18	JeniusConnect		78	82251	@pjharuu Untuk	0	0		bank btpn	99	99
14	15	2021-12-0	1,5E+18	JeniusConnect		78	82251	@sayangbgtyoha	0	0		bank btpn	99	99
15	17	2021-07-0	1,4E+18	Bukanbun	bandung	43	24	@JeniusConnect	0	0		bank btpn	99	99
16	18	2021-03-0	1,4E+18	bingungga	jakarta	2007	230	@JeniusConnect	0	0		bank btpn	-1	0
17	19	2021-01-1	1,3E+18	saekhanur	disini	656	7924	@yourmo0od @	0	0		bank btpn	99	99
18	20	2022-06-2	1,5E+18	larastika		291	386	@JeniusConnect	0	0		bank btpn	1	2
19	21	2022-02-1	1,5E+18	LuciferInC	Banten, In	357	247	@JeniusConnect	0	0		bank btpn	99	99
20	22	2021-04-1	1,4E+18	tamubiasaa		161	23	@JeniusConnect	0	0		bank btpn	1	2
21	23	2022-01-1	1,5E+18	chanthege	one piece	1855	1701	@JeniusConnect	0	0		bank btpn	-1	0
22	25	2021-02-2	1,4E+18	nonasoere	Jakarta	417	484	@AmadL @Jeniu	0	0		bank btpn	99	99
23	26	2021-07-2	1,4E+18	Tanyajago	Indonesia	2	3893	@achtzig_prozer	0	1		bank btpn	99	99
24	27	2021-02-1	1,4E+18	adeesulae	adeesulae	817	435	Ga tau kapan ma	0	0		bank btpn	-1	0
25	28	2021-08-2	1,4E+18	jeniushelp		2	27930	@luthfanar langs	0	0		bank btpn	99	99

Sheet1

Data yang Digunakan

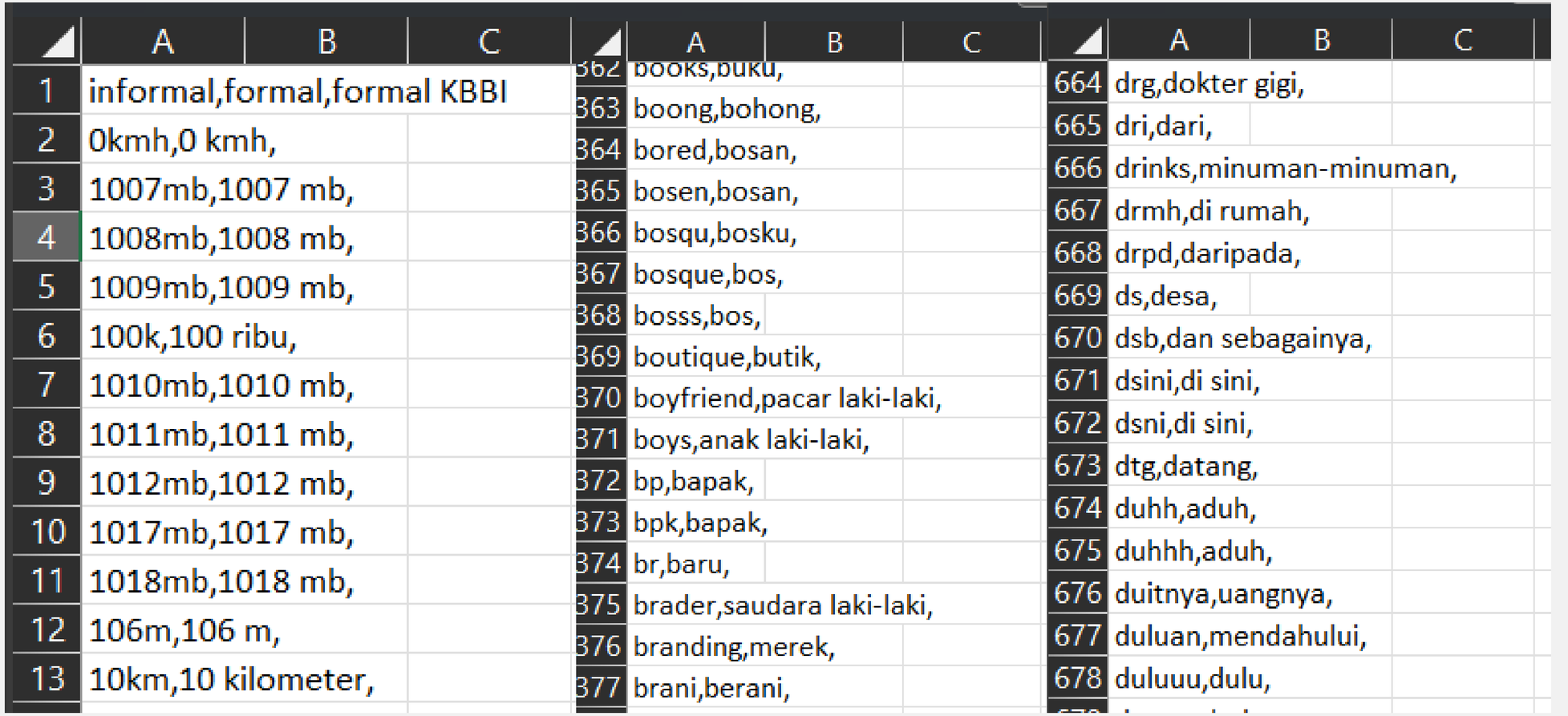
3. curr_train

	A	B	C	D	E	F
1	no	comments	bank	date	platform	Label (1,0,-1)
2	1	Keren banget Transmart 😍😍	Bank Mega	2023-05-07 00:00:00	Instagram	1
3	2	gapernah di kabarin apply cc	Bank Mega	2023-05-07 00:00:00	Instagram	-1
4	3	Keren bht	Bank Mega	2023-05-07 00:00:00	Instagram	1
5	4	Transmart di makssar masih kurang lengkap barangnya	Bank Mega	2023-05-07 00:00:00	Instagram	-1
6	5	@bankmegaid & Transmart keren Oke	Bank Mega	2023-05-07 00:00:00	Instagram	1
7	6	Transmart keren pokoknyaa	Bank Mega	2023-05-07 00:00:00	Instagram	1
8	7	@ayu.prima_ Keren bgt	Bank Mega	2023-05-07 00:00:00	Instagram	1
9	8	Sukses selalu semakin jaya 💖😍	Bank Mega	2023-05-07 00:00:00	Instagram	1
10	9	Biasanya ada yg komen "namaku alfan" tumben ga ada	Bank Mega	2023-05-07 00:00:00	Instagram	1
11	10	Mantappp Sukses selalu @bankmegaid	Bank Mega	2023-05-06 00:00:00	Instagram	1
12	11	sukses selalu 😍	Bank Mega	2023-05-06 00:00:00	Instagram	1
13	12	Terimakasih atas informasinya min 😍♥	Bank Mega	2023-05-07 00:00:00	Instagram	1
14	13	No 2 Sih Min	Bank Mega	2023-05-06 00:00:00	Instagram	1
15	14	makasih infonya	Bank Mega	2023-05-06 00:00:00	Instagram	1
16	15	Karna saya suka menghabiskan sebagian besar Uang saya untuk kebutuhan fashion,style mode terbaru untuk menunjang penampilan aja dan kalo misalkan lagi ada acara Selalu beli dan wajib beli hehehe Karana saya menghormati setiap moment yg ada...gitu sih Bank Mega salah satu Bank Kebanggaan Saya.	Bank Mega	2023-05-05 00:00:00	Instagram	1
	16	Susah kaya karna selalu boros dan selalu royal dalam pengeluaran apalagi nabung dibank yg biaya potongannya gede nasabah bukannya saldo jd nambah malah jadi saldo				1

Patrick | Vito | Farel | +

Data yang Digunakan

4. 230217 - Kamus NLP - slang_informal



	A	B	C
1	informal,formal,formal KBBI		
2	0kmh,0 kmh,		
3	1007mb,1007 mb,		
4	1008mb,1008 mb,		
5	1009mb,1009 mb,		
6	100k,100 ribu,		
7	1010mb,1010 mb,		
8	1011mb,1011 mb,		
9	1012mb,1012 mb,		
10	1017mb,1017 mb,		
11	1018mb,1018 mb,		
12	106m,106 m,		
13	10km,10 kilometer,		

	A	B	C
362	books,buku,		
363	boong,bohong,		
364	bored,bosan,		
365	bosen,bosan,		
366	bosqu,bosku,		
367	bosque,bos,		
368	bosss,bos,		
369	boutique,butik,		
370	boyfriend,pacar laki-laki,		
371	boys,anak laki-laki,		
372	bp,bapak,		
373	bpk,bapak,		
374	br,baru,		
375	brader,saudara laki-laki,		
376	branding,merek,		
377	brani,berani,		

	A	B	C
664	drg,dokter gigi,		
665	dri,dari,		
666	drinks,minuman-minuman,		
667	drmh,di rumah,		
668	drpd,daripada,		
669	ds,desa,		
670	dsb,dan sebagainya,		
671	dsini,di sini,		
672	dsni,di sini,		
673	dtg,datang,		
674	duhh,aduh,		
675	duhhh,aduh,		
676	duitnya,uangnya,		
677	duluan,mendahului,		
678	duluuu,dulu,		

Preprocessing & EDA

1. Import Library

```
import pandas as pd
import re
import os
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
import seaborn as sns
from wordcloud import WordCloud
import plotly.express as px
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.probability import FreqDist
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import torch
import emoji
from googletrans import Translator
from sklearn.model_selection import train_test_split
from tqdm import tqdm
import nltk
nltk.download('stopwords')
```

Preprocessing & EDA

2. Read excel data to pandas dataframe

```
df_raw_p = pd.read_excel("curr_train.xlsx",'Patrick')
df_raw_v = pd.read_excel("curr_train.xlsx",'Vito')
df_raw_f = pd.read_excel("curr_train.xlsx",'Farel')
df_mod_mega = pd.read_excel("df_sentiment_modified.xlsx",'mega')
df_mod_btpn = pd.read_excel("df_sentiment_modified.xlsx",'btpn')
df_mod_btn = pd.read_excel("df_sentiment_modified.xlsx",'btn')
df_all = pd.read_excel("df_all.xlsx",'Sheet1')
```

Jumlah Data :

df_raw_p : 19995

df_raw_v : 6634

df_raw_f : 18979

df_mod_mega : 4654

df_mod_btpn : 9359

df_mod_btn : 5996

df_all : 19777

3. Modify Data

```
df_mod_mega['bank'] = 'Bank Mega'
df_mod_btpn['bank'] = 'BTPN'
df_mod_btn['bank'] = 'BTN'
```

```
df_all['platform'] = 'Instagram'
df_mod_mega['platform'] = 'Instagram'
df_mod_btpn['platform'] = 'Instagram'
df_mod_btn['platform'] = 'Instagram'
```

```
[29] df_mod_mega.rename(columns={'text': 'comments','sentiment':'Label (1,0,-1)'}, inplace=True)
     df_mod_btpn.rename(columns={'text': 'comments','sentiment':'Label (1,0,-1)'}, inplace=True)
     df_mod_btn.rename(columns={'text': 'comments','sentiment':'Label (1,0,-1)'}, inplace=True)
     df_all.rename(columns={'text': 'comments','sentiment_ori':'Label (1,0,-1)','category':'bank','u
```

Drop unused column

```
df_raw_p = df_raw_p.drop(['no','date'], axis=1)
df_raw_v = df_raw_v.drop(['no','date'], axis=1)
df_raw_f = df_raw_f.drop(['no','date'], axis=1)
df_mod_mega = df_mod_mega.drop(['user','sentiment_mod'], axis=1)
df_mod_btpn = df_mod_btpn.drop(['user','sentiment_mod'], axis=1)
df_mod_btn = df_mod_btn.drop(['user','sentiment_mod'], axis=1)
df_all = df_all[['comments','Label (1,0,-1)','bank']]
```

Preprocessing & EDA

4. Handle Missing Value

```
[26] df_raw_p = df_raw_p[df_raw_p['Label (1,0,-1)'].notna()]
     df_raw_p = df_raw_p[df_raw_p['comments'].notna()]
     df_raw_p
```

```
[27] df_raw_v = df_raw_v[df_raw_v['Label (1,0,-1)'].notna()]
     df_raw_v = df_raw_v[df_raw_v['comments'].notna()]
     df_raw_v
```

```
df_raw_f = df_raw_f[df_raw_f['Label (1,0,-1)'].notna()]
df_raw_f = df_raw_f[df_raw_f['comments'].notna()]
df_raw_f
```

```
[30] df_mod_mega = df_mod_mega[df_mod_mega['Label (1,0,-1)'].notna()]
     df_mod_mega = df_mod_mega[df_mod_mega['comments'].notna()]
     df_mod_mega
```

Jumlah Data :

df_raw_p :19995

df_raw_v : 6634

df_raw_f : 7809

df_mod_mega : 4654

df_mod_btpn : 9359

df_mod_btn : 5996

df_all : 19777

Preprocessing & EDA

5. Combine Dataframe

```
[39] df_raw = pd.concat([df_raw_p,df_raw_v,df_raw_f,df_mod_mega,df_mod_btpn,df_mod_btn,df_all])
     df_raw
```

Jumlah Data Total: 74224

6. Change unwanted value

Change unwanted value

```
df_raw['Label (1,0,-1)'].unique()
```

```
array([1, -1, 0, 99, 9, 11, -11, '-'], dtype=object)
```

```
[41] df_raw.loc[df_raw['Label (1,0,-1)'] == 9, 'Label (1,0,-1)'] = 99 #ubah typo 9 menjadi 99
     df_raw.loc[df_raw['Label (1,0,-1)'] == 11, 'Label (1,0,-1)'] = 1 #ubah typo 11 menjadi 1
     df_raw.loc[df_raw['Label (1,0,-1)'] == -11, 'Label (1,0,-1)'] = -1 #ubah typo -11 menjadi -1
     df_raw.loc[df_raw['Label (1,0,-1)'] == '-', 'Label (1,0,-1)'] = -1 #ubah typo - menjadi -1
```

```
[42] df_raw['Label (1,0,-1)'].unique()
```

```
array([1, -1, 0, 99], dtype=object)
```

Label sudah sesuai yang diinginkan dimana 1 adalah komentar positif, 0 netral, -1 negatif, dan 99 sebagai data yang duplikat. Maka selanjutnya data dengan label 99 akan dihapus

Preprocessing & EDA

7. Erase data with 99 label

Erase data with 99 label

```
[43] df_raw = df_raw[df_raw['Label (1,0,-1)'] != 99]
```

0s

```
[44] df_raw['Label (1,0,-1)'].unique()
```

0s

```
array([1, -1, 0], dtype=object)
```

Jumlah Data :
74224 -> 55057

8. Change Data Type

```
df_raw['comments'] = df_raw['comments'].astype(str)
df_raw['bank'] = df_raw['bank'].astype(str)
df_raw['platform'] = df_raw['platform'].astype(str)
df_raw['Label (1,0,-1)'] = df_raw['Label (1,0,-1)'].astype(int)
```

Preprocessing & EDA

9. Format Bank Name

Format Bank Name

```
[47] df_raw['bank'].unique()
```

```
array(['Bank Mega', 'Allo Bank', 'Neo Bank', 'BRI', 'BCA', 'BTPN', 'BTN',
       'bank btpn', 'bank mega', 'bank btn'], dtype=object)
```

```
[48] df_raw.loc[df_raw['bank'] == 'bank mega', 'bank'] = 'Bank Mega'
     df_raw.loc[df_raw['bank'] == 'mega', 'bank'] = 'Bank Mega'
     df_raw.loc[df_raw['bank'] == 'btpn', 'bank'] = 'Bank BTPN'
     df_raw.loc[df_raw['bank'] == 'bank btpn', 'bank'] = 'Bank BTPN'
     df_raw.loc[df_raw['bank'] == 'btn', 'bank'] = 'Bank BTN'
     df_raw.loc[df_raw['bank'] == 'bank btn', 'bank'] = 'Bank BTN'
     df_raw.loc[df_raw['bank'] == 'BTN', 'bank'] = 'Bank BTN'
     df_raw.loc[df_raw['bank'] == 'BTPN', 'bank'] = 'Bank BTPN'
     df_raw.loc[df_raw['platform'] == 'facebook', 'platform'] = 'Facebook'
```

```
[49] df_raw['bank'].unique()
```

```
array(['Bank Mega', 'Allo Bank', 'Neo Bank', 'BRI', 'BCA', 'Bank BTPN',
       'Bank BTN'], dtype=object)
```

Preprocessing & EDA

10. Remove words with @

Remove words with @

```
df_raw['text_cleaned'] = df_raw['comments'].str.replace(r'@\w+\b', '', regex=True)
df_raw
```

11. Remove words with

Remove words with "#"

```
df_raw['text_cleaned'] = df_raw['text_cleaned'].str.replace(r'#\w+\b', '')
df_raw['text_cleaned']
```

12. Remove words with https

Remove words with "https"

```
df_raw['text_cleaned'] = df_raw['text_cleaned'].str.replace(r'https\S*', '')
df_raw
```

Preprocessing & EDA

13. Remove duplicates & Empty comments

Remove duplicates & empty comments

```
[54] df_raw.drop_duplicates(inplace=True)
     df_raw
```

```
[55] df_raw = df_raw[df_raw['text_cleaned']!='']
     df_raw
```

Jumlah Data setelah data duplikat dihapus : 42309

Jumlah Data setelah data kosong dihapus : 41726

Preprocessing & EDA

14. Change Slang Data

Change Slang Data

```
[56] slang_words_1 = {"@": "di", "abis": "habis", "ad": "ada", "adlh": "adalah", "afaik": "as far as i know", "ahaha": "haha", "aj": "saja",
                      "ajep-ajep": "dunia gemerlap", "ak": "saya", "akika": "aku", "akkoh": "aku", "akuwh": "aku", "alay": "norak", "alow": "halo",
                      "ambilin": "ambilkan", "ancur": "hancur", "anjrit": "anjing", "anter": "antar", "ap2": "apa-apa", "apasih": "apa sih", "apes":
                      "aps": "apa", "aq": "saya", "aquwh": "aku", "asbun": "asal bunyi", "aseekk": "asyik", "asekk": "asyik", "asem": "asam",
                      "aspal": "asli tetapi palsu", "astul": "asal tulis", "ato": "atau", "au ah": "tidak mau tahu", "awak": "saya", "ay": "sayang",
                      "ayank": "sayang", "b4": "sebelum", "bakalan": "akan", "bandes": "bantuan desa", "bangedh": "sangat", "banpol": "bantuan polis
                      "banpur": "bantuan tempur", "basbang": "basi", "bcanda": "bercanda", "bdg": "bandung", "begajulan": "nakal", "beliin": "belika
```

```
slang_words2 = pd.read_csv('230217 - Kamus NLP - slang_informal.csv', header=0)
slang_words2 = slang_words2[['informal', 'formal']]
slang_words_2 = dict(slang_words2.values)
```

```
[58] slang_words = dict(slang_words_1.items() | slang_words_2.items())
```

```
[59] list_sentence_train = []
     for sentence in tqdm(df_raw['text_cleaned']) :
         cleaned_sentence = [slang_words[word] if word in list(slang_words.keys()) else word for word in str(sentence).split()]
         list_sentence_train.append(' '.join(cleaned_sentence))
     df_raw['review_text_cleaned'] = list_sentence_train
```

```
100%|██████████| 41726/41726 [00:51<00:00, 816.32it/s]
```

Kamus kata slang didapat melalui file 230217 - Kamus NLP - slang_informal.csv dan kamus yang berada di gdrive

Preprocessing & EDA

15. Train Test Split

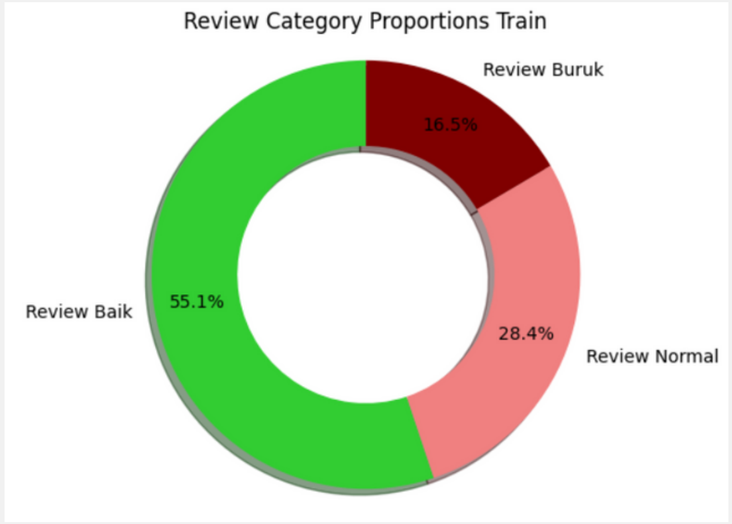
Train Test Split

```
[60] df_train, df_test = train_test_split(df_raw, test_size=0.2, random_state=42)
```

0s

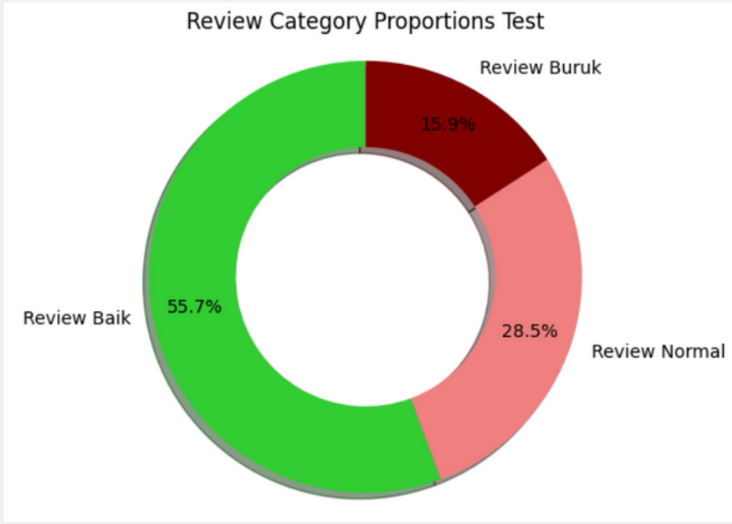
16. Data Label Proportion

- Data Train



Positif : 18402
Netral : 9466
Negatif : 5512

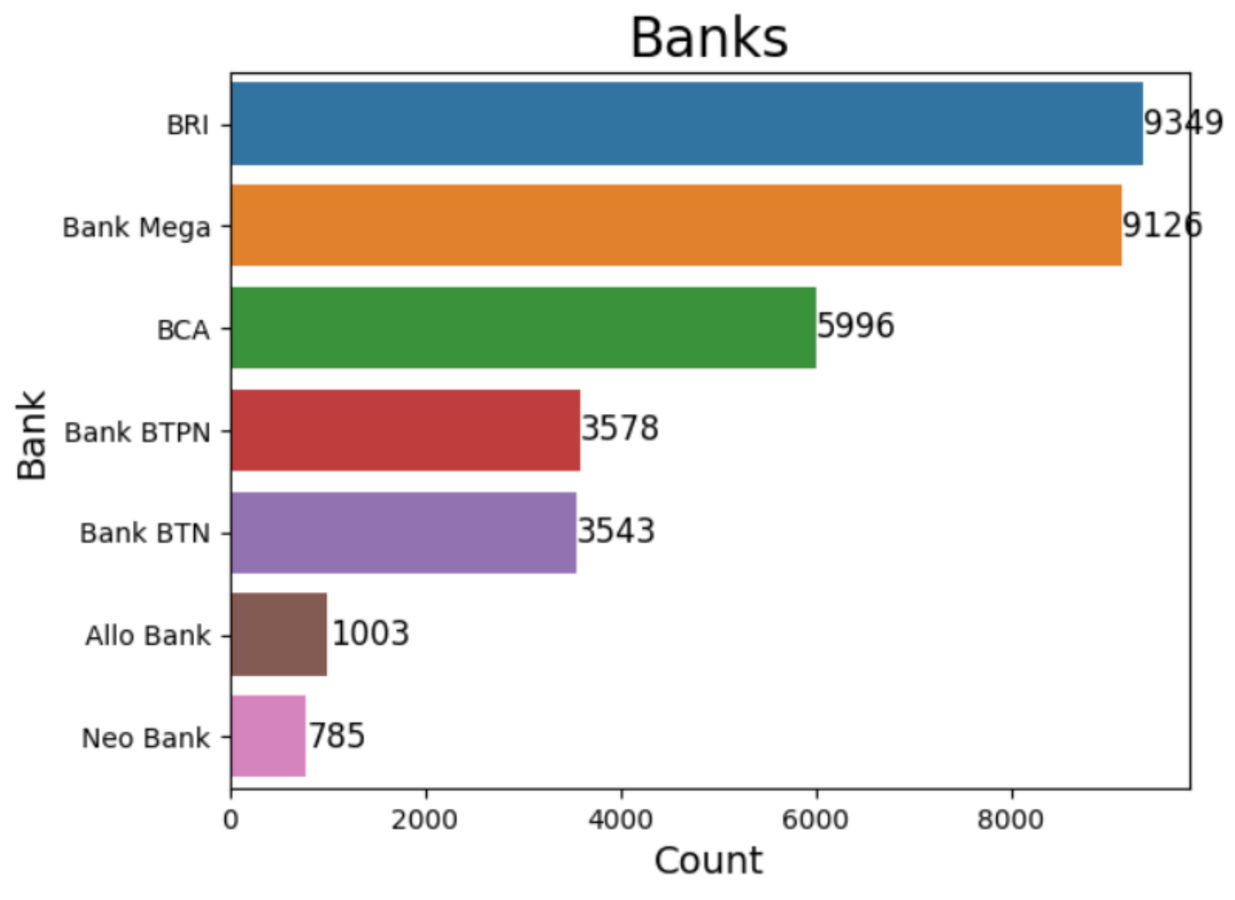
- Data Test



Positif : 4647
Netral : 2375
Negatif : 1324

Preprocessing & EDA

17. Bank Accounts



BRI : 9349

Bank Mega : 9126

BCA : 5996

Bank BTPN : 3578

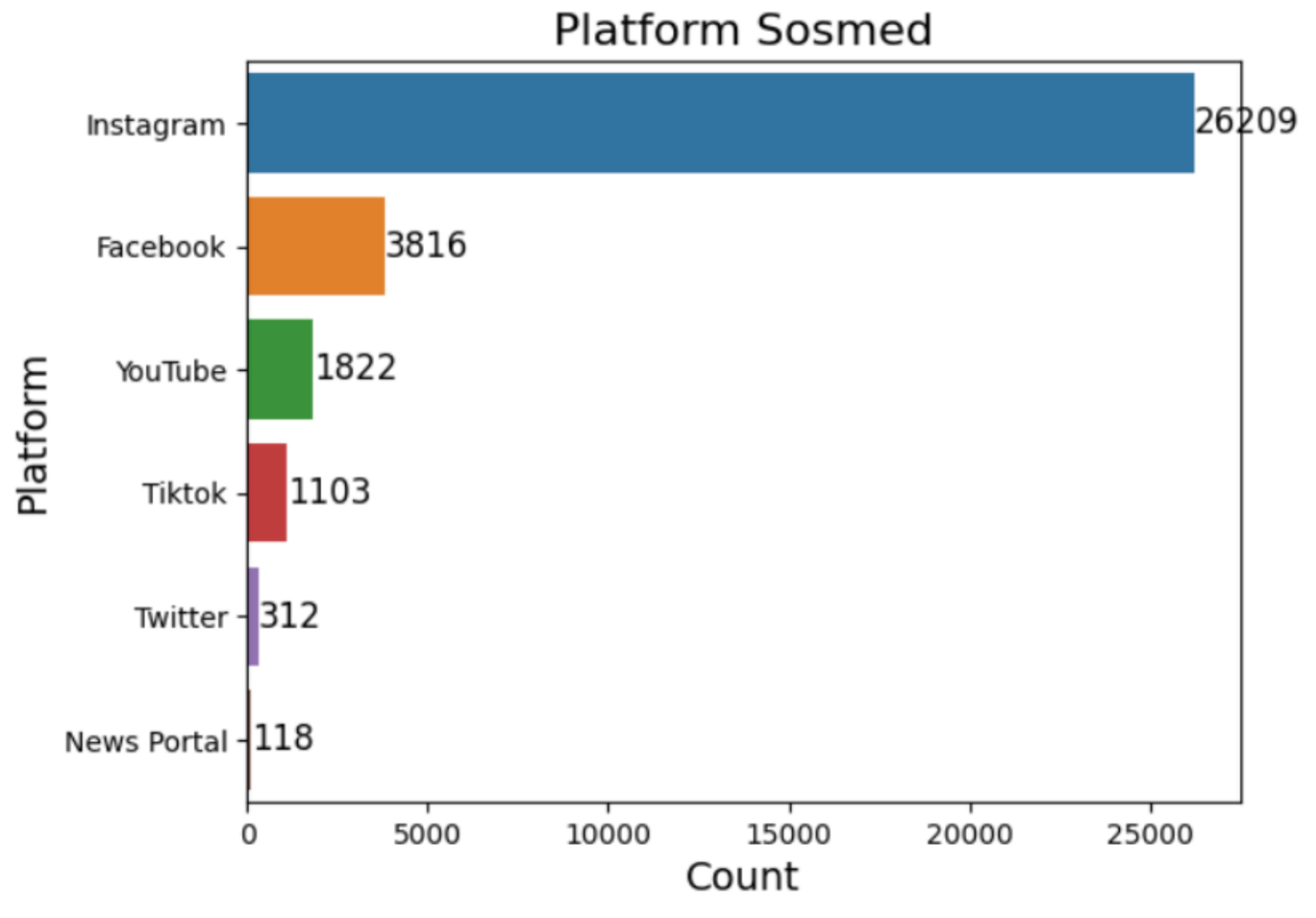
Bank BTN : 3543

Allo Bank : 1003

Neo Bank : 785

Preprocessing & EDA

18. Social Media Platform



Instagram : 26209

Facebook : 3816

YouTube : 1822

Tiktok : 1103

Twitter : 312

News Portal : 118

Preprocessing & EDA

19. Review Dengan Emoji

```
comments                @dagangkorea 536,340 bca/dana/btpn  -🦊
bank                                                Bank BTPN
platform                                            Instagram
Label (1,0,-1)                                              0
text_cleaned                      536,340 bca/dana/btpn  -🦊
review_text_cleaned                536,340 bca/dana/btpn -🦊
Name: 4022, dtype: object
comments                Ada banyak diskon 😍
bank                               Bank Mega
platform                           Instagram
Label (1,0,-1)                             1
text_cleaned            Ada banyak diskon 😍
review_text_cleaned     Ada banyak diskon 😍
Name: 7302, dtype: object
```

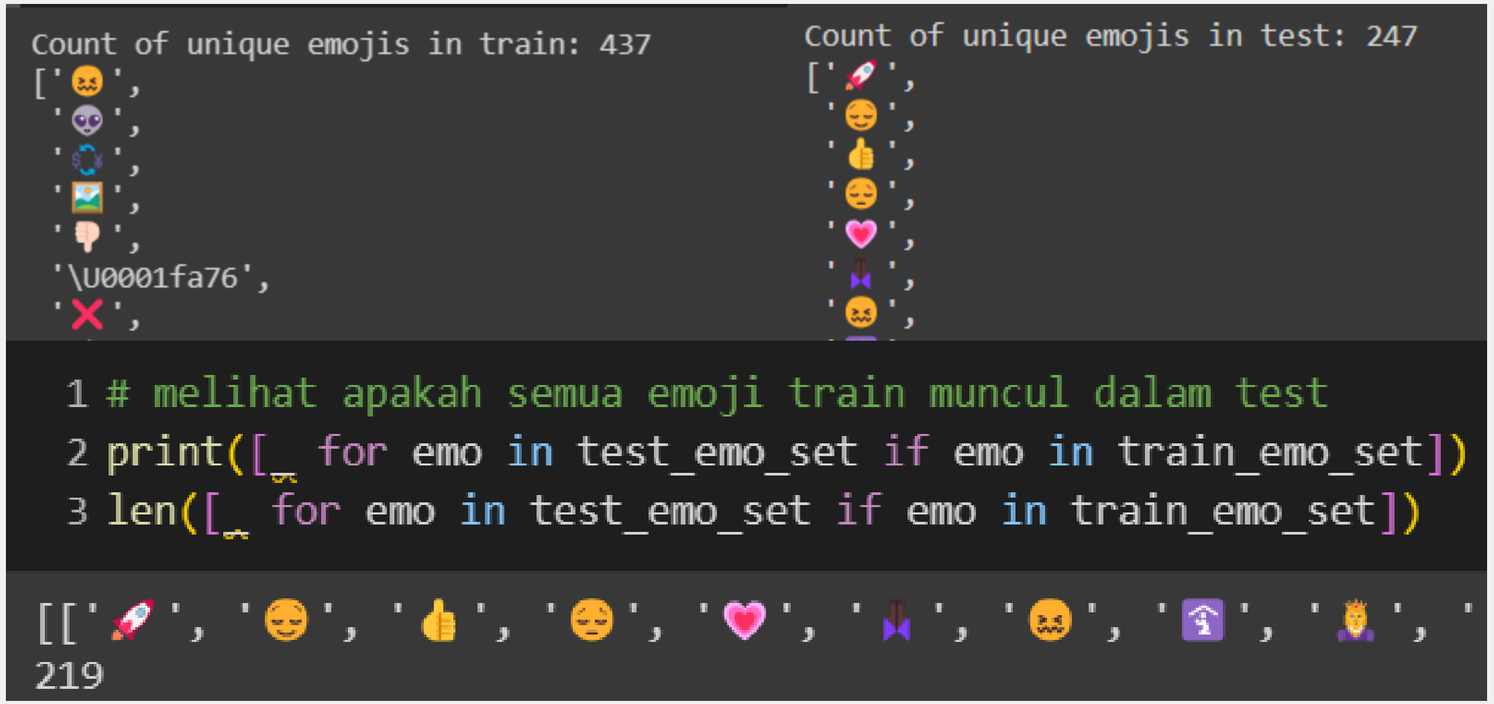
- Terdapat 8696 data train yang memiliki emoji dari 33847 data
- Terdapat 2155 data test yang memiliki emoji dari 8462 data

20. Demojize

```
'536,340 bca/dana/btpn -:fox:'
'Ada banyak diskon :smiling_face_with_heart-eyes:'
```

Preprocessing & EDA

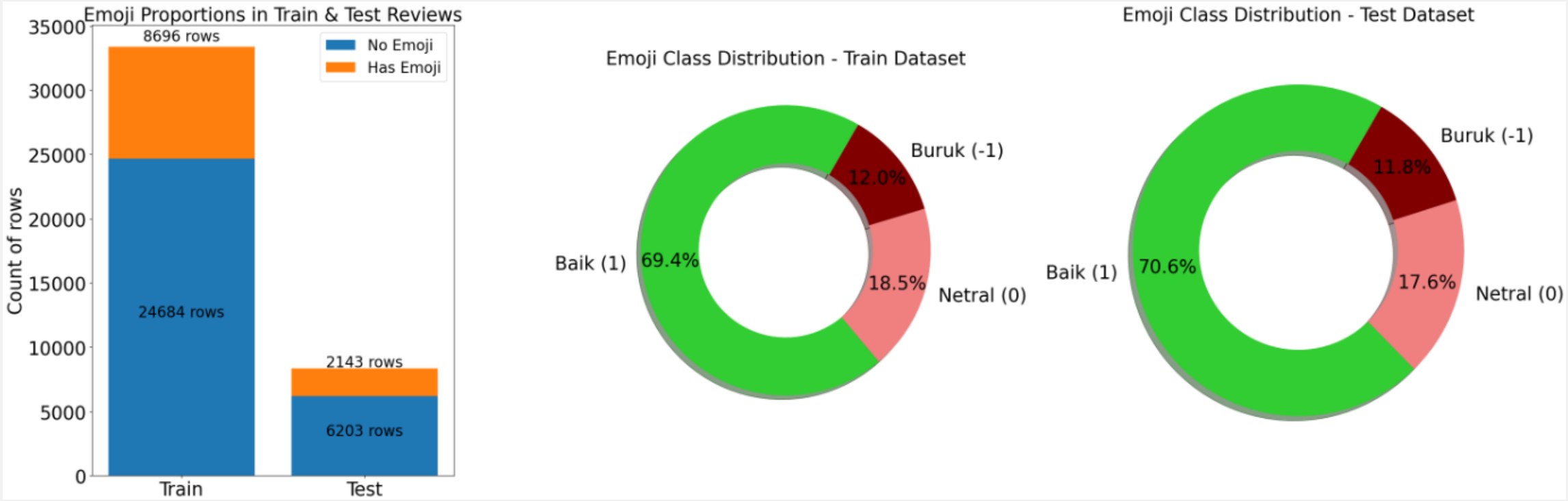
21. Unique Emoji



Terdapat 247 - 219 = 28 emoji yang terdapat pada test namun tidak ada dalam train

Preprocessing & EDA

21. Sentimen Komentar ber-Emoji



- Terdapat 8696 data train ber-emoji dan 24684 tidak ber-emoji.

- Terdapat 2143 data test ber-emoji dan 6203 tidak ber-emoji.

Train Data ber-emoji:

Positif = 6038

Netral = 1612

Negatif = 1046

Test Data ber-emoji:

Positif = 1514

Netral = 377

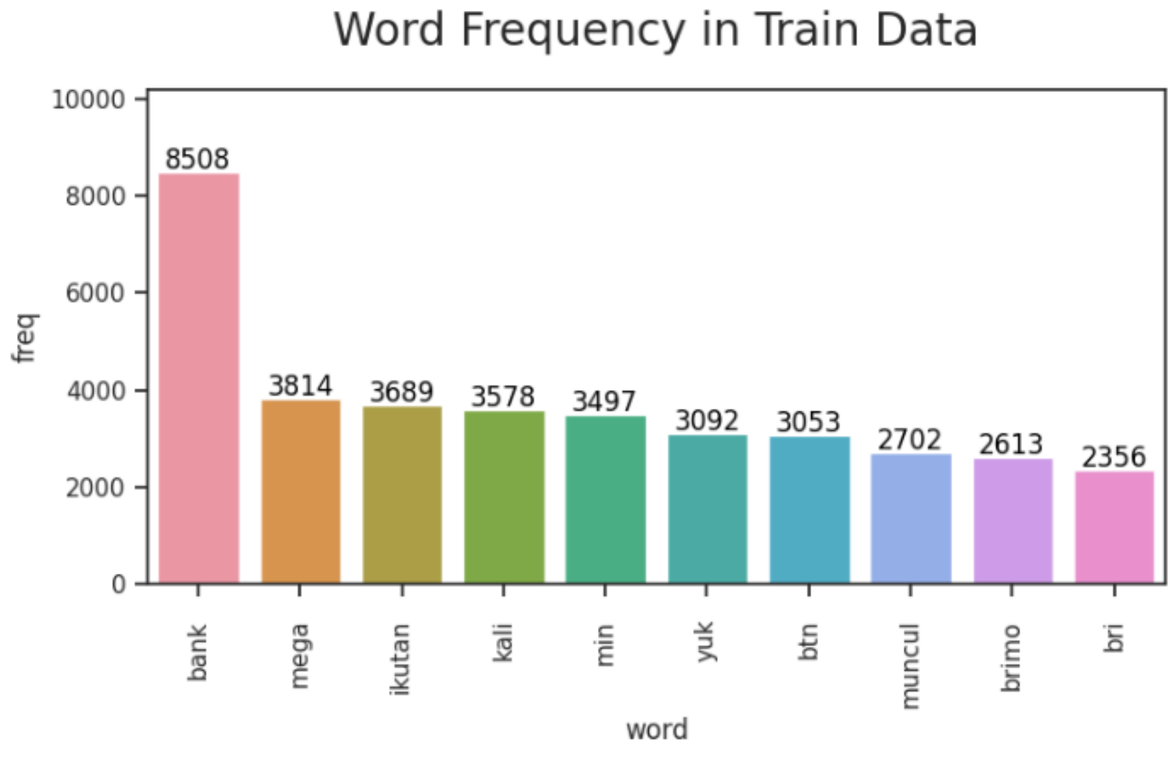
Negatif = 252

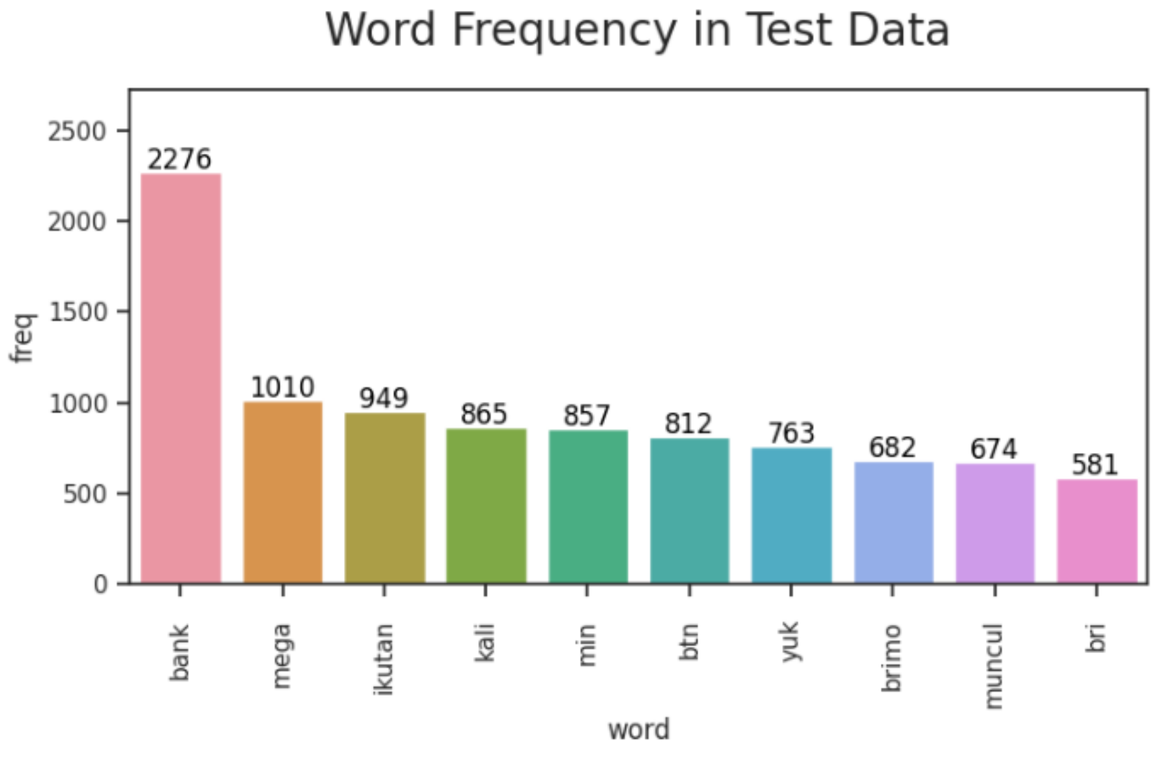
Preprocessing & EDA

22. Corpus Kata

```
Count of unique words in train: 26436
Count of unique words in test: 11617
```

Terdapat 26461 kata unik dalam Data Train dan 11526 pada data Test





Preprocessing & EDA

23. Word Cloud

- Data Train

- Data Test

Preprocessing & EDA

24. Clean and Demojize Data

- Memberi kurung pada emoji

```
# 1. Memberi kutung pada emoji
def add_brackets_to_emoji(text):
    emojified_text = emoji.emojize(text)
    modified_text = ""
    for char in emojified_text:
        if emoji.is_emoji(char):
            modified_text += " [" + char + "] "
        else:
            modified_text += char
    return modified_text
```

- Full clean

```
# 2. Full clean
def full_clean(text):
    text = add_brackets_to_emoji(text)
    # print(text)
    # demojize first
    text = emoji.demojize(text,delimiters=("", ""))

    # lower text
    cleaned = re.sub('[^a-zA-Z0-9\[\]]+',' ',text).lower()

    # remove 2 letter words
    shortword = re.compile(r'\W*\b\w{1,2}\b')
    cleaned = re.sub(shortword, '', cleaned)

    # double whitespace to single
    cleaned = re.sub('[ ]+',' ',cleaned)

    return cleaned
```

Preprocessing & EDA

25. Menghapus Data kosong setelah dibersihkan

```
# mencari data yang kosong setelah dibersihkan
df_empty_after_cleaned_train = df_train['Data'].loc[(df_train['Data']=='')|(df_train['Data']==' ')].index
len(df_empty_after_cleaned_train)
```

```
203
```

Terdapat 203 data komentar kosogn pada data train. Setelah dibersihkan jumlah data train menjadi 33177

26. Panjang kata pada data

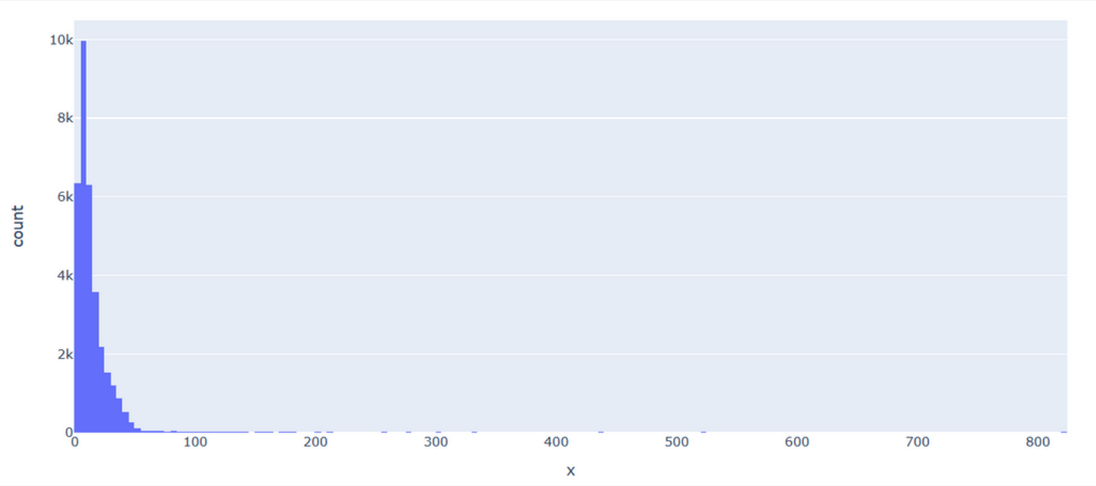
```
Rekapitulasi dari panjang ulasan:

- Train -
Max =  822
Min =  1
Average =  13.449498146306176
StdDev =  13.92562847762983

- Test -
Max =  269
Min =  0
Average =  13.290079079798707
StdDev =  12.710950494230845
```

	comments	bank	platform	Label (1,0,-1)	text_cleaned	review_text_cleaned	emoji	Data	tokenized	length_of_review
30179	Jujur gua sebenarnya ga mau komentar, tpi ga e...	Bank Mega	Facebook	0	Jujur gua sebenarnya ga mau komentar, tpi ga e...	Jujur saya sebenarnya tidak mau komentar, teta...	no emoji	jujur saya sebenarnya tidak mau komentar tetap...	[jujur, saya, sebenarnya, tidak, mau, komentar...	822

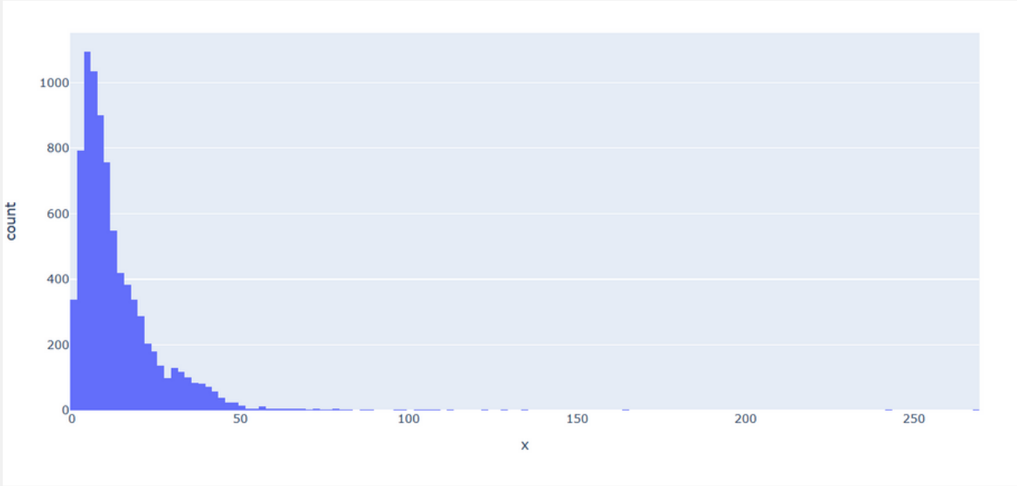
Preprocessing & EDA

27. Distribusi panjang kata pada data

- Data Train



- Data Test



Preprocessing & EDA

28. Translate Data

```python
def find_en(df):
    result = []
    datas = df['Data'].reset_index(drop=True)
    # print(datas)
    # print(len(datas))
    for i in tqdm(range(len(datas))):
        curr_word = ''
        # print(i)
        # print(datas[i])
        for word in re.split(r'\s+(?![^[\]]*\])', datas[i]):
            word = word.strip('[]')
            # print(word)
            if word != '' and word != None:
                if translator.detect(word) != None and translator.detect(word).lang == 'en':
                    curr_word += ' '
                    curr_word += trans_to_id(word)
                else:
                    curr_word += ' '
                    curr_word += word
        result.append(curr_word)
        # print(curr_word)
    df['translated'] = result
    return df
```

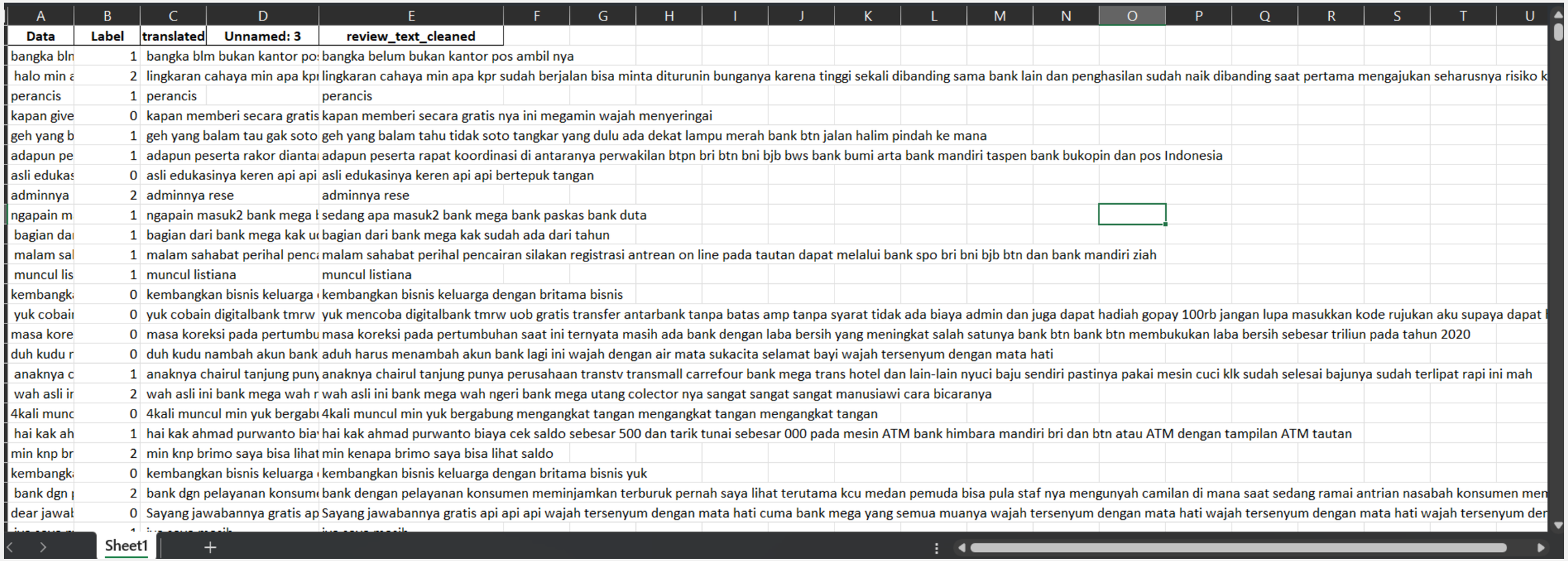
Masing-masing kata akan di iterasi dan kata-kata dengan bahasa inggris akan diterjemahkan kedalam bahasa Indonesia menggunakan API Google Translate.

Kata-kata yang merupakan emoji yang telah diubah menjadi bahasa inggris, akan diterjemahkan secara bersamaan dengan memanfaatkan kurung kotak “[]”.

Hasil Data yang telah diterjemahkan akan disimpan pada file df_train_fix.xlsx dan df_test_fix.xlsx untuk selanjutnya digunakan dalam pembuatan model.

Penyesuaian Data

Karena Terdapat data-data yang memiliki label tidak sesuai, maka data dilakukan labeling ulang dan beberapa data dihapus. Sehingga data final yang digunakan dalam pelatihan model adalah data fix_fixed.xlsx



Data	Label	translated	Unnamed: 3	review_text_cleaned
bangka bln	1	bangka blm bukan kantor po		bangka belum bukan kantor pos ambil nya
halo min a	2	lingkaran cahaya min apa kp		lingkaran cahaya min apa kpr sudah berjalan bisa minta diturunin bunganya karena tinggi sekali dibanding sama bank lain dan penghasilan sudah naik dibanding saat pertama mengajukan seharusnya risiko k
perancis	1	perancis		perancis
kapan give	0	kapan memberi secara gratis		kapan memberi secara gratis nya ini megamin wajah menyeringai
geh yang b	1	geh yang balam tau gak soto		geh yang balam tahu tidak soto tangkar yang dulu ada dekat lampu merah bank btn jalan halim pindah ke mana
adapun pe	1	adapun peserta rakor dianta		adapun peserta rapat koordinasi di antaranya perwakilan btpn bri btn bni bjb bws bank bumi arta bank mandiri taspen bank bukopin dan pos Indonesia
asli edukas	0	asli edukasinya keren api api		asli edukasinya keren api api bertepuk tangan
adminnya	2	adminnya rese		adminnya rese
ngapain m	1	ngapain masuk2 bank mega b		sedang apa masuk2 bank mega bank paskas bank duta
bagian dar	1	bagian dari bank mega kak u		bagian dari bank mega kak sudah ada dari tahun
malam sal	1	malam sahabat perihal penc		malam sahabat perihal pencairan silakan registrasi antrean on line pada tautan dapat melalui bank spo bri bni bjb btn dan bank mandiri ziah
muncul lis	1	muncul listiana		muncul listiana
kembangk	0	kembangkan bisnis keluarga		kembangkan bisnis keluarga dengan britama bisnis
yuk cobair	0	yuk cobain digitalbank tmrw		yuk mencoba digitalbank tmrw uob gratis transfer antarbank tanpa batas amp tanpa syarat tidak ada biaya admin dan juga dapat hadiah gopay 100rb jangan lupa masukkan kode rujukan aku supaya dapat
masa kore	0	masa koreksi pada pertumbu		masa koreksi pada pertumbuhan saat ini ternyata masih ada bank dengan laba bersih yang meningkat salah satunya bank btn bank btn membukukan laba bersih sebesar triliun pada tahun 2020
duh kudu r	0	duh kudu nambah akun bank		aduh harus menambah akun bank lagi ini wajah dengan air mata sukacita selamat bayi wajah tersenyum dengan mata hati
anaknya c	1	anaknya chairul tanjung puny		anaknya chairul tanjung punya perusahaan transtv transmall carrefour bank mega trans hotel dan lain-lain nyuci baju sendiri pastinya pakai mesin cuci klk sudah selesai bajunya sudah terlipat rapi ini mah
wah asli ir	2	wah asli ini bank mega wah r		wah asli ini bank mega wah ngeri bank mega utang colector nya sangat sangat sangat manusiawi cara bicaranya
4kali munc	0	4kali muncul min yuk bergab		4kali muncul min yuk bergabung mengangkat tangan mengangkat tangan mengangkat tangan
hai kak ah	1	hai kak ahmad purwanto bia		hai kak ahmad purwanto biaya cek saldo sebesar 500 dan tarik tunai sebesar 000 pada mesin ATM bank himbara mandiri bri dan btn atau ATM dengan tampilan ATM tautan
min knp br	2	min knp brimo saya bisa lihat		min kenapa brimo saya bisa lihat saldo
kembangk	0	kembangkan bisnis keluarga		kembangkan bisnis keluarga dengan britama bisnis yuk
bank dgn p	2	bank dgn pelayanan konsum		bank dengan pelayanan konsumen meminjamkan terburuk pernah saya lihat terutama kcu medan pemuda bisa pula staf nya mengunyah camilan di mana saat sedang ramai antrian nasabah konsumen men
dear jawal	0	Sayang jawabannya gratis ap		Sayang jawabannya gratis api api api wajah tersenyum dengan mata hati cuma bank mega yang semua muanya wajah tersenyum dengan mata hati wajah tersenyum dengan mata hati wajah tersenyum der

Naive Bayes

- Tokenizer

```
from sklearn.feature_extraction.text import CountVectorizer
from nltk.tokenize import RegexpTokenizer
token = RegexpTokenizer(r'[a-zA-Z0-9]+')
cv = CountVectorizer(ngram_range = (1,1),tokenizer = token.tokenize)
text_counts = cv.fit_transform(combined_df['translated'])
```

- ComplementNB

```
ComplementNB model accuracy is 72.63%
-----------------------------------------------
Confusion Matrix:
      0     1     2
0  1104   206    87
1   536  1427   407
2   411   623  3494
-----------------------------------------------
Classification Report:
              precision    recall  f1-score   support

          -1       0.54      0.79      0.64      1397
           0       0.63      0.60      0.62      2370
           1       0.88      0.77      0.82      4528

    accuracy                           0.73      8295
...
    accuracy                           0.73      8295
   macro avg       0.68      0.72      0.69      8295
weighted avg       0.75      0.73      0.73      8295
```

Naive Bayes

- MultinominalNB

```
MultinominalNB model accuracy is 74.08%
------------------------------------------------
Confusion Matrix:
      0     1     2
0  1067   225   105
1   386  1488   496
2   229   709  3590
------------------------------------------------
Classification Report:
              precision    recall  f1-score   support

          -1       0.63      0.76      0.69      1397
           0       0.61      0.63      0.62      2370
           1       0.86      0.79      0.82      4528

    accuracy                           0.74      8295
...
    accuracy                           0.74      8295
   macro avg       0.70      0.73      0.71      8295
weighted avg       0.75      0.74      0.74      8295
```

- BernoulliNB

```
BernoulliNB model accuracy = 72.74%
------------------------------------------------
Confusion Matrix:
     0     1     2
0  805   263   329
1  235  1314   821
2  107   506  3915
------------------------------------------------
Classification Report:
              precision    recall  f1-score   support

          -1       0.70      0.58      0.63      1397
           0       0.63      0.55      0.59      2370
           1       0.77      0.86      0.82      4528

    accuracy                           0.73      8295
...
    accuracy                           0.73      8295
   macro avg       0.70      0.67      0.68      8295
weighted avg       0.72      0.73      0.72      8295
```

LSTM Model 1

- Model

```
Model: "model_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_2 (InputLayer)        [(None, 1)]               0

 embedding_3 (Embedding)     (None, 1, 64)             640000

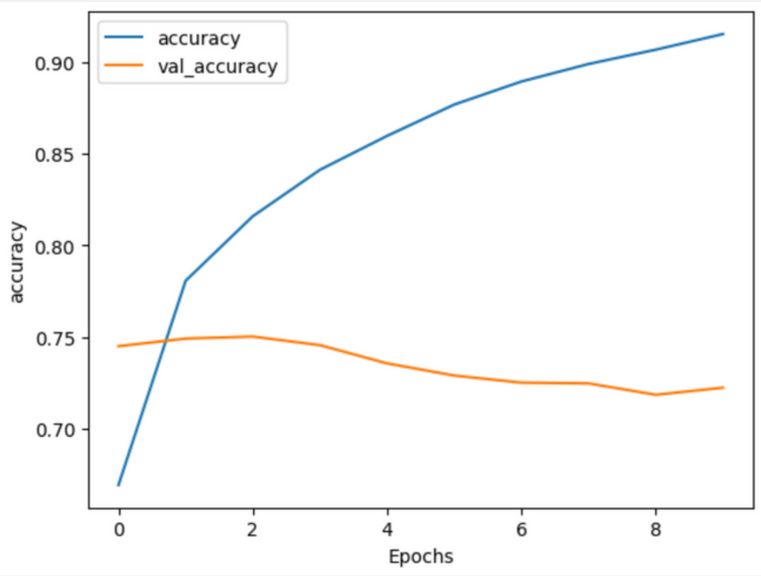
 bidirectional_5 (Bidirecti  (None, 128)               66048
 onal)

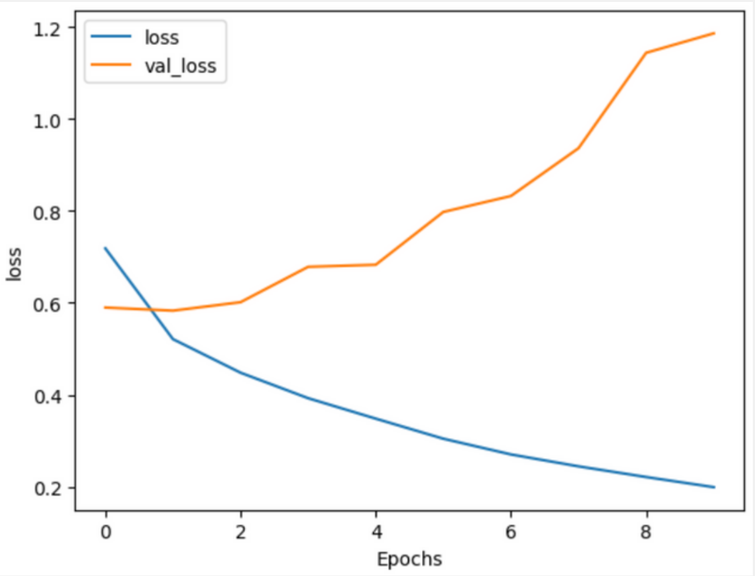
 dense_10 (Dense)            (None, 64)                8256

 dense_11 (Dense)            (None, 3)                 195

=================================================================
Total params: 714499 (2.73 MB)
...
Total params: 714499 (2.73 MB)
Trainable params: 714499 (2.73 MB)
Non-trainable params: 0 (0.00 Byte)
```

- Result





loss: 0.4986

accuracy: 0.7960

val_loss: 0.5533

val_accuracy: 0.7699

LSTM Model 2

- Model

```
=================================================================
embedding (Embedding)        (None, 100, 64)           8192

bidirectional (Bidirection   (None, 100, 128)          66048
al)

bidirectional_1 (Bidirecti   (None, 32)                18560
onal)

dense (Dense)                (None, 16)                528

leaky_re_lu (LeakyReLU)      (None, 16)                0

dropout (Dropout)            (None, 16)                0

batch_normalization (Batch   (None, 16)                64
Normalization)

dense_1 (Dense)              (None, 16)                272

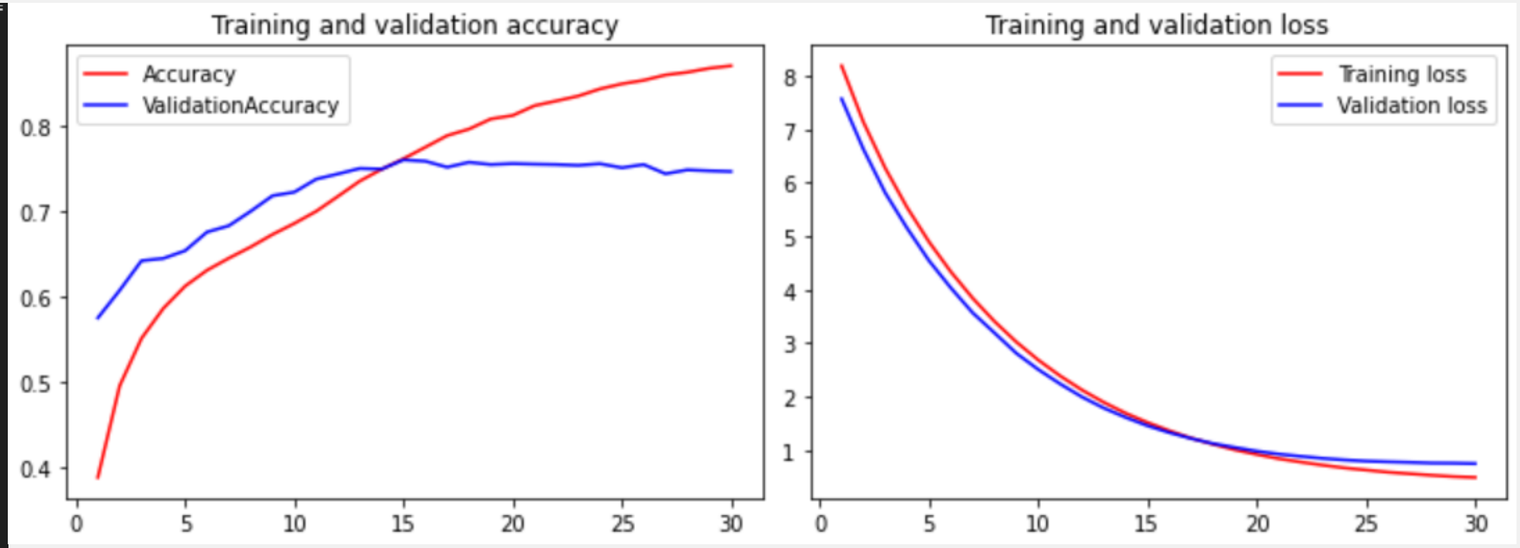
dropout_1 (Dropout)          (None, 16)                0

batch_normalization_1 (Bat   (None, 16)                64
chNormalization)

dense_2 (Dense)              (None, 8)                 136

dense_3 (Dense)              (None, 3)                 27
```

- Result



loss: 0.4844

accuracy: 0.8702

val_loss: 0.7427

val_accuracy: 0.7467

GRU

- Model

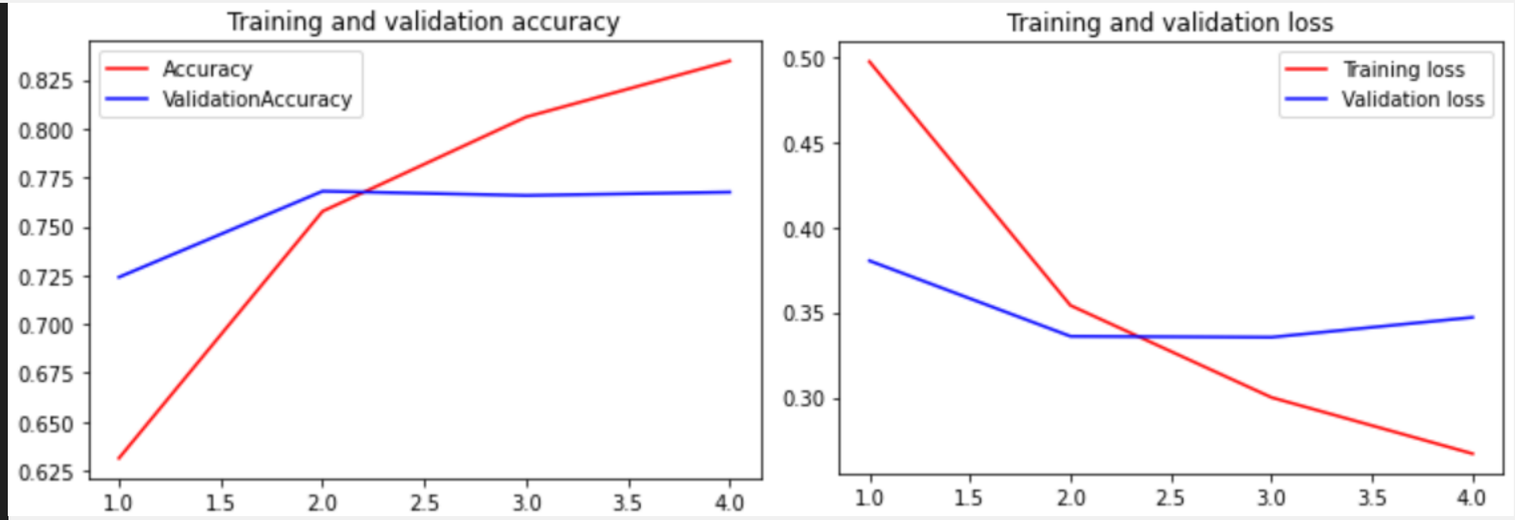
```
Build model...
Summary of the built model...
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_3 (Embedding)     (None, 15, 100)           1000000

 gru (GRU)                   (None, 32)                12864

 dense_10 (Dense)            (None, 3)                 99

=================================================================
Total params: 1,012,963
Trainable params: 1,012,963
...
Trainable params: 1,012,963
Non-trainable params: 0
```

- Result



loss: 0.2670

accuracy: 0.8346

val_loss: 0.3470

val_accuracy: 0.7675

IndoBERT - NaiveBayes

- Model

indobenchmark/indobert-large-p2

- Result Accuracy

MultinomialNB : 0.6443661971830986
BernoulliNB : 0.6443661971830986
ComplementNB : 0.2112676056338028

Fine-Tune IndoBERT

- Model

ayameRushia/bert-base-indonesian-1.5G-sentiment-analysis-smsa

Train_loss : 0.547500
Val_loss : 0.507889
Accuracy : 0.793438
Precision : 0.791858
Recall : 0.793438
F1 : 0.792387

- Result Accuracy

Step	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
50	1.312100	0.676318	0.713750	0.693103	0.713750	0.688276
100	0.691900	0.627476	0.739844	0.727589	0.739844	0.726266
150	0.624400	0.602664	0.743906	0.751269	0.743906	0.741497
200	0.611200	0.566809	0.765625	0.757350	0.765625	0.757085
250	0.615000	0.575515	0.762031	0.773424	0.762031	0.766298
300	0.603000	0.553771	0.771563	0.763210	0.771563	0.758572
350	0.577500	0.569431	0.760000	0.764834	0.760000	0.761917
400	0.582700	0.554137	0.767031	0.781654	0.767031	0.772069
450	0.596300	0.562172	0.765156	0.769286	0.765156	0.763125
500	0.618700	0.523946	0.779062	0.770847	0.779062	0.765573
550	0.524800	0.523569	0.780312	0.772515	0.780312	0.769673
600	0.605000	0.525606	0.773281	0.765449	0.773281	0.763533
650	0.533000	0.536822	0.776875	0.789778	0.776875	0.780120
700	0.566200	0.521178	0.782813	0.778316	0.782813	0.774726
750	0.552000	0.565325	0.777969	0.769763	0.777969	0.766093
800	0.552300	0.525813	0.783750	0.776148	0.783750	0.771366
850	0.553700	0.507117	0.787813	0.779757	0.787813	0.779290
900	0.511600	0.516863	0.785625	0.784906	0.785625	0.785224
950	0.552100	0.506054	0.795781	0.788210	0.795781	0.786724

Fine-Tune RoBERTA

- Model
ayameRushia/roberta-base-indonesian-1.5G-sentiment-analysis-smsa

- Result Accuracy

Validation Loss Epoch: 0.4981026891068905
Validation Accuracy Epoch: 79.53125
Accuracy on test data = 79.53%

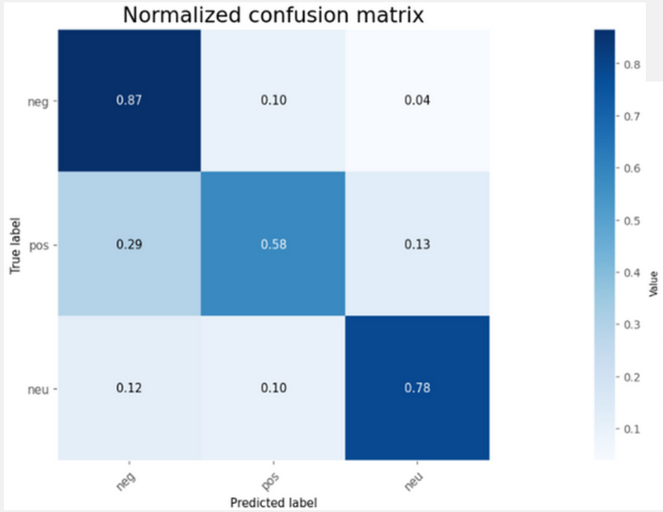
Fine-Tune GPT 2

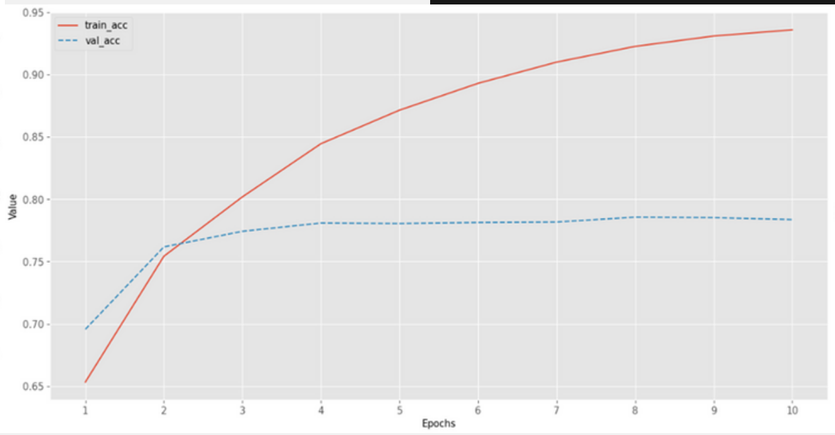
- Model

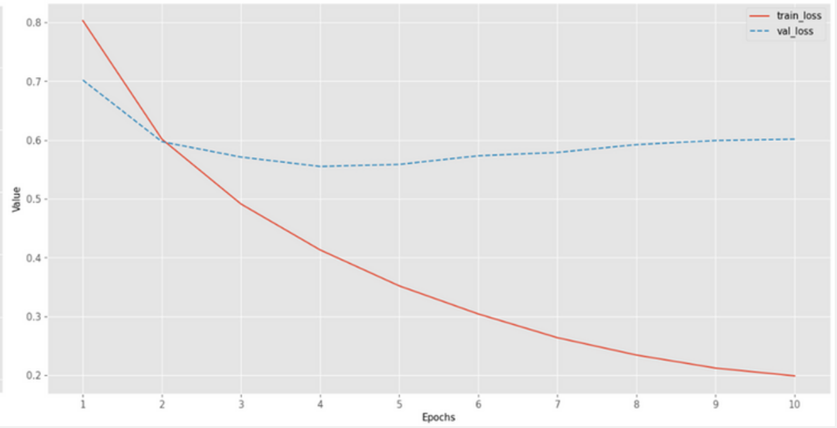
indonesian-nlp/gpt2

- Result Accuracy

```
model_name_or_path = 'indonesian-nlp/gpt2' #0.78599%
model_name_or_path = 'indonesian-nlp/gpt2-medium-indonesian' #0.78482%
model_name_or_path = 'cahya/gpt2-large-indonesian-522M' #kebesaran
model_name_or_path = 'flax-community/gpt2-medium-indonesian' #0.78208%
model_name_or_path = 'cahya/gpt2-medium-indonesian' #76%
model_name_or_path = 'indonesian-nlp/gpt2-medium-indonesian' #78%
model_name_or_path = 'cahya/gpt2-small-indonesian-522M' #78%
```







Prompt Engineering LLaMA 2

- Prompt

Determine the comment sentiment polarity in the given review delimited by triple quotes.

Determine the sentiment polarity from the options [\"positive\", \"negative\", \"neutral\"].

Only answer with a format of [\"sentiment\"] without any explanation.

If the polarity can't be determined just answer [\"neutral\"].

```{text}```

- Result Accuracy

  Accuracy : 37.25%

- Example

```
coba = 'kali muncul logo brimo'
print(llm_chain.run(coba))
```

```
...
 Neutral.
```
```

Thank You!