

Sentiment Analysis TikTok Apps Based On Google Play Store Reviews





MIND MAP

01

USE CASE

02

CRISP-DM

03

SUMMARY

PROBLEM & SOLUTION



PROBLEM

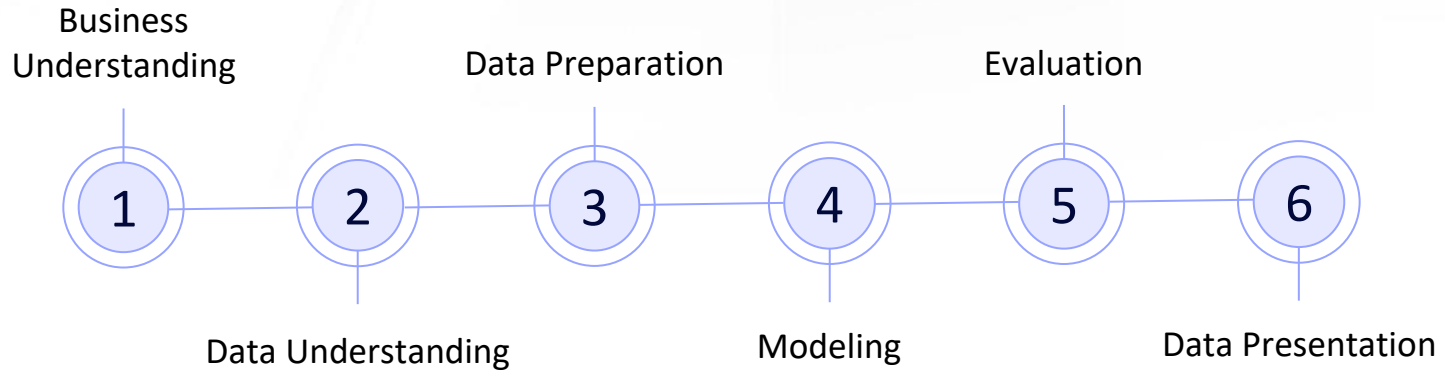
Bagaimana menganalisa sentiment publik ketika menggunakan aplikasi tiktok berdasarkan review di google play store?



SOLUTION

Dengan *machine learning* memungkinkan untuk menganalisa sentiment publik terhadap penggunaan suatu aplikasi apakah tergolong positif, negative atau netral.

CRISP-DM





BUSINESS UNDERSTANDING

TikTok adalah sebuah aplikasi jejaring sosial dan platform *video music* dimana pengguna bisa membuat, mengedit, dan berbagi klip video pendek lengkap dengan filter dan disertai musik sebagai pendukung.

Dengan aplikasi TikTok, pengguna dapat membuat video pendek yang unik dengan cepat dan juga mudah untuk dibagikan dengan teman dan ke seluruh dunia.



DATA UNDERSTANDING

Data didapat dengan melakukan scraping di website Google Play Store untuk mengambil review dari aplikasi tiktok.

`df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 686691 entries, 0 to 686690
Data columns (total 11 columns):
#   Column              Non-Null Count  Dtype  
---  -
0   Unnamed: 0           686691 non-null int64  
1   reviewId             686691 non-null object 
2   userName             686689 non-null object 
3   userImage            686691 non-null object 
4   content              686685 non-null object 
5   score               686691 non-null int64  
6   thumbsUpCount        686691 non-null int64  
7   reviewCreatedVersion 377004 non-null object 
8   at                   686691 non-null object 
9   replyContent         6734 non-null  object 
10  repliedAt            6734 non-null  object 
dtypes: int64(3), object(8)
memory usage: 57.6+ MB
```

`df.head()`

	Unnamed: 0	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt
0	0	gp:AOqpTOFENoHk7-Xp_qYh_rQIXXy9_Io4YA7PKweeQ9...	Syamil Mohammad	https://play-lh.googleusercontent.com/a-/AOh14...	The best	5	0	20.8.4	2021-08-29 21:40:38	NaN	NaN
1	1	gp:AOqpTOEQITicaQfuanys5AmIFM0NJXHpE7I0v6YRbyE...	Jx Tuason	https://play-lh.googleusercontent.com/a/AATXAJ...	This App Is Cool You Will Watch Tutorial Or Dance	5	0	NaN	2021-08-29 21:37:03	NaN	NaN
2	2	gp:AOqpTOETcrjZdrDoymAwjy_xy6VuXLaYqdoH5BpdzQ7...	Jacky hng	https://play-lh.googleusercontent.com/a-/AOh14...	Love it so much	5	0	20.8.4	2021-08-29 21:35:40	NaN	NaN
3	3	gp:AOqpTOEF0S4QNB99BEISTCQHTzy4hulrFounZtBOuQL...	Aron Carpio	https://play-lh.googleusercontent.com/a-/AOh14...	ldk man	2	0	NaN	2021-08-29 21:35:02	NaN	NaN
4	4	gp:AOqpTOFTczKfPSKZmE6sp3GvNqA3ETGvgKJKPV3Dall...	xxz gurl	https://play-lh.googleusercontent.com/a/AATXAJ...	@shlxcxn	5	0	NaN	2021-08-29 21:34:59	NaN	NaN

DATA PREPARATION



DATA PREPARATION

Menyiapkan library yang diperlukan seperti **pandas**, **seaborn**, **sklearn** dan dataset.



DATA CLEANSING

Drop kolom yang tidak dipakai, menghapus string/ emoji yang tidak dipakai, mengatur format penulisan kata.



EXPLORATORY DATA ANALYSIS

Melihat distribusi tiap sentiment untuk mengambil insight.

DATA PREPARATION

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 686691 entries, 0 to 686690
Data columns (total 11 columns):
#   Column              Non-Null Count  Dtype
---  ---
0   Unnamed: 0           686691 non-null  int64
1   reviewId             686691 non-null  object
2   userName             686689 non-null  object
3   userImage            686691 non-null  object
4   content              686685 non-null  object
5   score                686691 non-null  int64
6   thumbsUpCount        686691 non-null  int64
7   reviewCreatedVersion 377004 non-null  object
8   at                   686691 non-null  object
9   replyContent         6734 non-null   object
10  repliedAt            6734 non-null   object
dtypes: int64(3), object(8)
memory usage: 57.6+ MB
```

drop

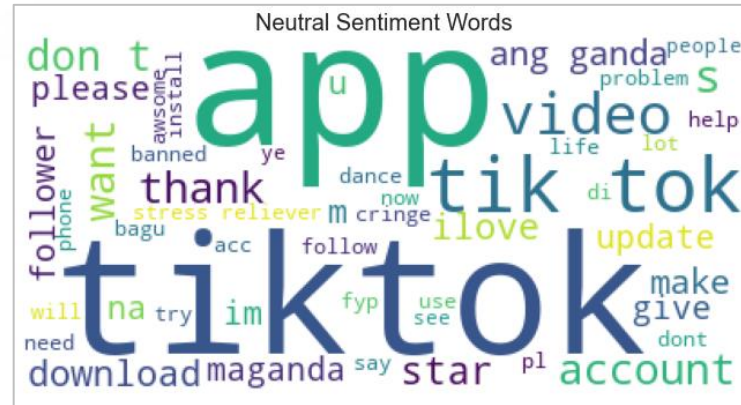
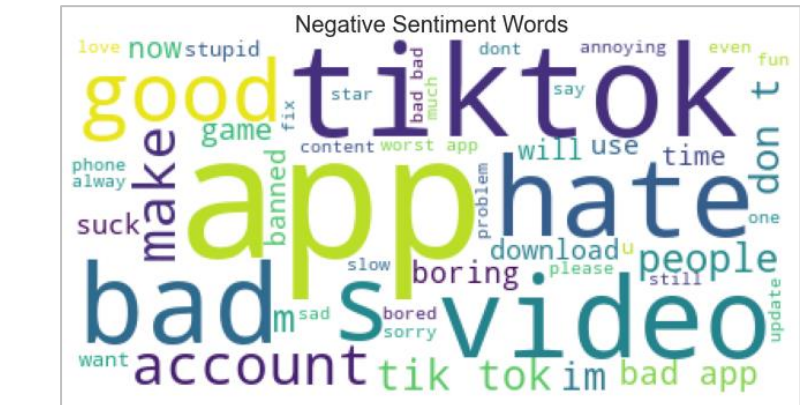
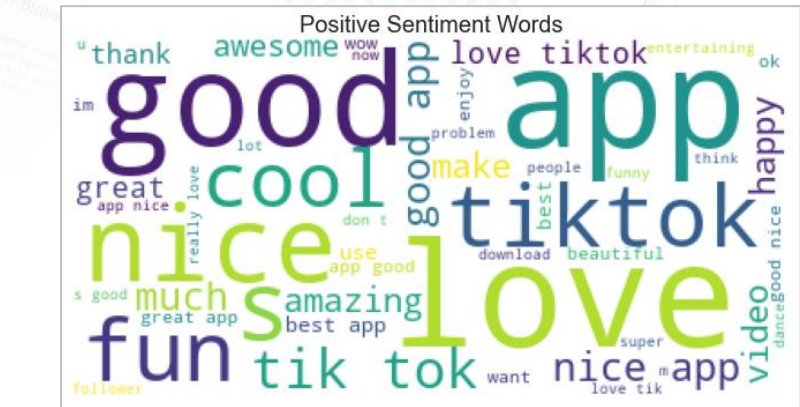
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 686691 entries, 0 to 686690
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  ---
0   content 686685 non-null  object
1   score   686691 non-null  int64
dtypes: int64(1), object(1)
memory usage: 10.5+ MB
```

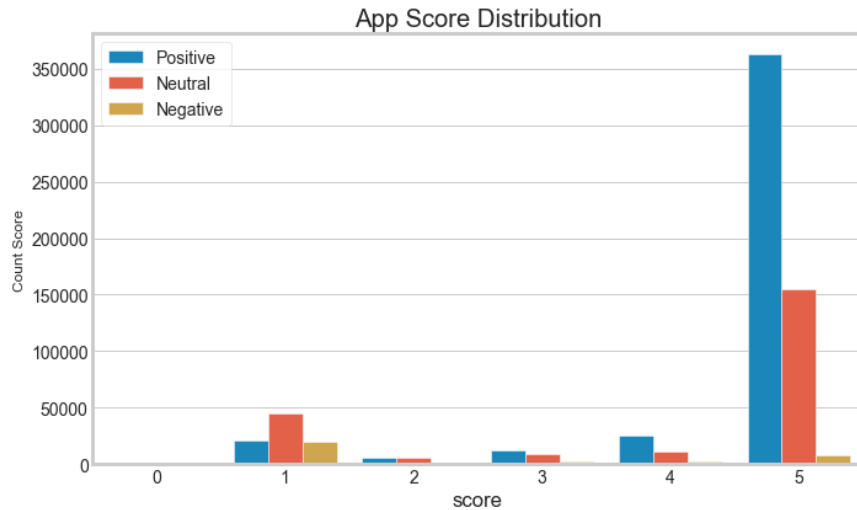
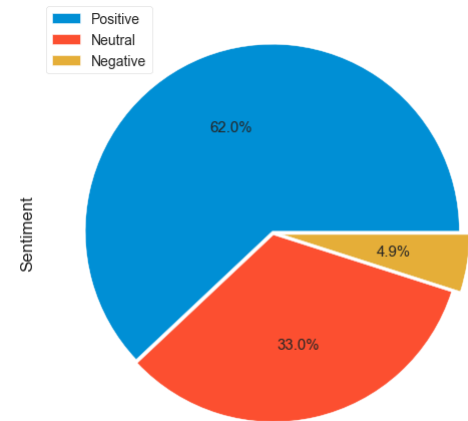
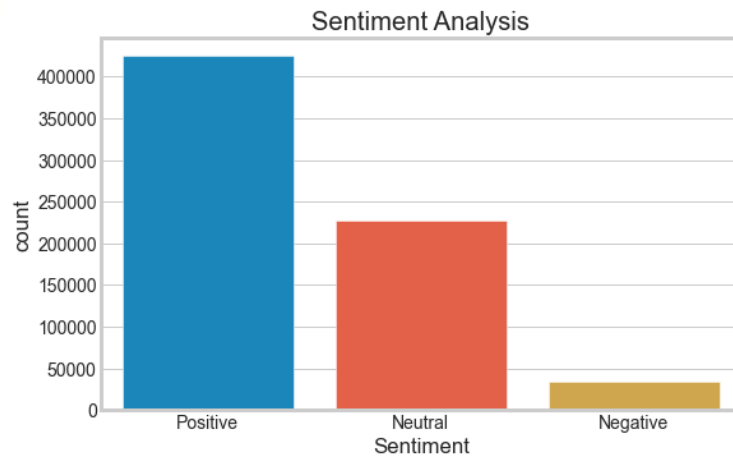
formatting

	content	score
0	The best	5
1	This App Is Cool You Will Watch Tutorial Or Dance	5
2	Love it so much	5
3	idk man	2
4	@shicxn	5

	content	score
0	the best	5
1	this app is cool you will watch tutorial or dance	5
2	love it so much	5
3	idk man	2
4	shicxn	5

A decorative graphic featuring three overlapping circles in blue, yellow, and grey on the left. On the right, there is a bar chart comparing data for 2017/18 (red bars) and 2016/17 (blue bars). The chart shows an increase in the 2017/18 data for most categories. A pen is visible on the far right.

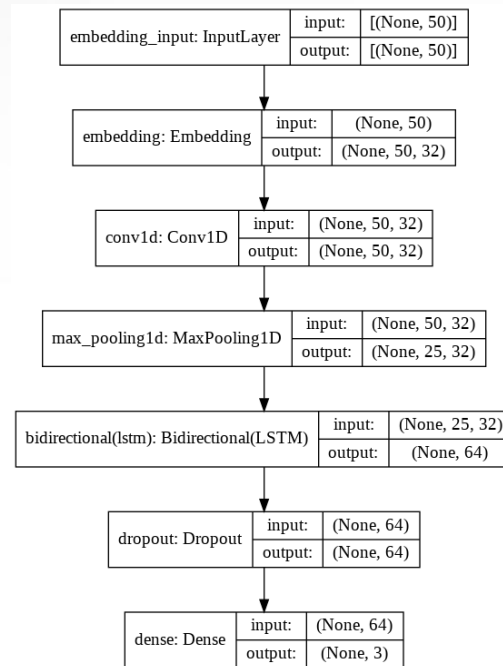




MODELING

Modeling Bidirectional LSTM dengan Neural Network dengan properties :

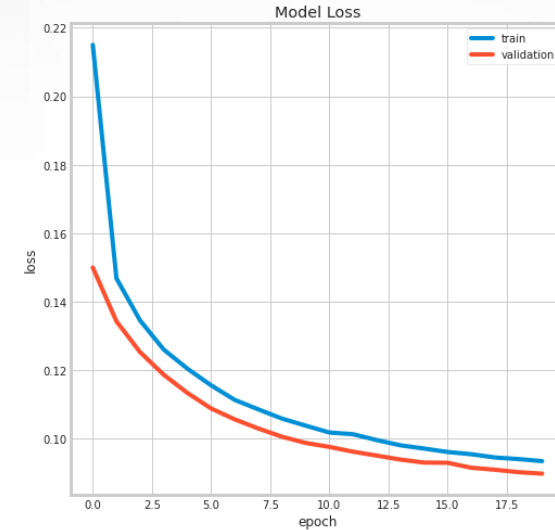
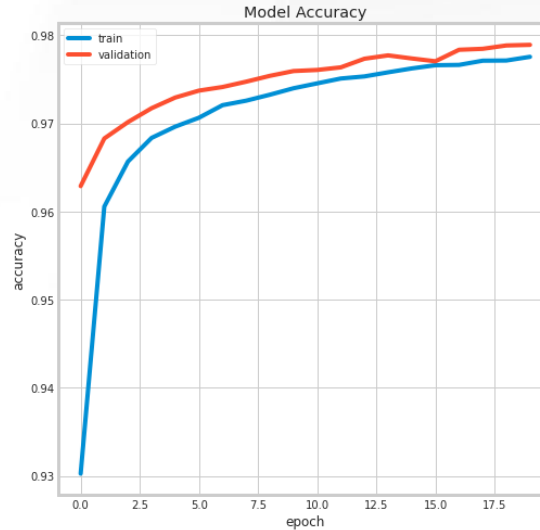
```
vocab_size = 5000  
embedding_size = 32  
epochs=20  
learning_rate = 0.1  
decay_rate = learning_rate / epochs  
momentum = 0.8
```



EVALUATION

Hasil evaluasi model menggunakan
LSTM Neural Network:

- Accuracy : 0.9792
- Precision : 0.9799
- Recall : 0.9781
- F1 Score : 0.9790





MODEL TEST

Kita mencoba menyimpan hasil model dan mengaplikasikan pada suatu review dan hasilnya sebagai berikut

```
predict_class(['this app is cool you will watch tutorial or dance'])
```

The predicted sentiment is Positive



```
predict_class(['so hard to use'])
```

The predicted sentiment is Negative



```
predict_class(['book'])
```

The predicted sentiment is Neutral





SUMMARY

- Dengan machine learning memungkinkan untuk mengetahui sentiment penggunaan aplikasi berdasarkan review di google playstore.
- Data reviews di google playstore banyak yang mengandung emoji atau kode ASCII lain sehingga ketika dilakukan *formatting data* banyak unsur yang hilang sehingga masuk di sentiment netral.
- Dari data yang diambil, aplikasi tiktok memiliki 62% sentiment positif, hal ini menunjukkan bahwa tiktok cukup banyak disukai oleh publik.
- Dengan menggunakan modeling LSTM Neural Network mendapatkan hasil performa yang cukup bagus.
- Dari hasil uji coba model untuk menganalisa sentiment, terlihat bahwa model berjalan dengan baik.
- Dari pihak developer aplikasi TikTok khususnya akan terbantu dengan adanya sentiment analisis ini mengingat perlu adanya pembaharuan aplikasi secara berkala.