

Data Pipeline Anteraja Review

TextBlob-based Sentiment Analysis

Our Team



Hakim
440307



Adhitya
444033



Hafidz
444052



Samatha
444071

Background



Platform *e-commerce* menggunakan strategi promo di tanggal cantik.

(Aria, 2020)



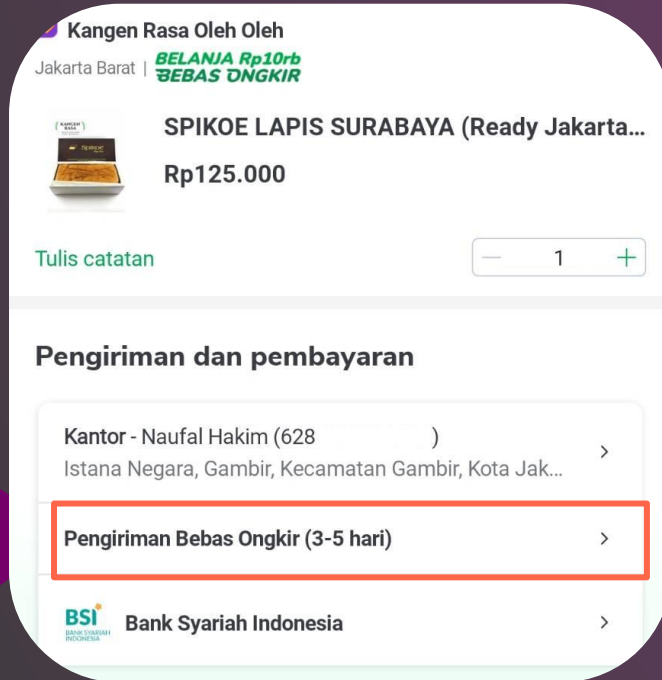
Terjadi peningkatan aktivitas logistik yang signifikan.

(Asosiasi Logistik Indonesia, 2019)



Diperlukan *insight* apakah peningkatan aktivitas logistik mempengaruhi sentimen publik terhadap perusahaan *low-cost logistic*.

Kenapa Anteraja?



Anteraja melayani pengiriman **gratis ongkir**
(Subsidi pengiriman Rp20.000 tetapi dengan
durasi pengiriman yang lebih lama).

ETL Process

Extract



Play Store
Review



Twitter Review
(Tweepy)

PySpark

Spark
Dataframe

Transform



Google
Translate API

PySpark

udf_clean

PySpark

udf_sentiment
_score (TextBlob)

PySpark

udf_sentiment

Load



AWS S3

Visualization



Seaborn

Table of contents

01

Extract

Get user reviews of Anteraja app
from Twitter & Google Play Store

02

Transform

Data cleaning and translation to
get well-defined data

03

Load

Loading the transformed data
into data warehouse

04

Visualization

Create
sentiment analysis charts



01

Extract

Get user reviews of Anteraja app
from Twitter & Google Play Store

Twitter + Play Store Review

```
search_words = "anteraja -from:anteraja_id lang:id"
date_year = "2021"
date_month = "10"
date_days = range(10, 18)
date_time = "2315"

tweets = tw.Cursor(api.search_full_archive,
                    environment_name="TwitterResearch",
                    query=search_words,
                    fromDate=date_since,
                    toDate=date_until).items(100)
```

Tweepy

```
id_reviews = reviews_all(
    'id.anteraja.aca',
    sleep_milliseconds=0,
    lang='en',
    country='id',
    sort=Sort.NEWEST,
)
```

google-play-scraper

02

Transform

Data cleaning and translation to get well-defined data. Also getting sentiment score from each data.



Transform

Translate

Translate reviews from
Twitter and Google Play

Get Sentiment Score

Get sentiment scores
from each review

1

2

3

4

Clean

Remove unnecessary
signs so that only user
reviews are left

Get Sentiment

Get sentiment category
for every review based
on the sentiment scores

translator()

```
def translator(text):  
    text = GoogleTranslator(source='id',  
                             target='en').translate(text)  
    return text
```

Translating from lang:id to lang:en with
Google Translator API

udf_clean

```
def cleaner(text):  
    text = re.sub("@[A-Za-z0-9]+", "", text) #Remove @ sign  
    text = ''.join(c for c in text if c not in emoji.UNICODE_EMOJI) #Remove Emojis  
    text = text.replace("#", "").replace("_", "") #Remove hashtag sign but keep the text  
    text = " ".join(w for w in nltk.wordpunct_tokenize(text) \  
        if w.lower() in words or not w.isalpha()) #Remove non-english tweets (not 100% success)  
    return text  
  
udf_clean = udf(lambda x:cleaner(x), StringType())
```

Clean data from @username, emoji, #Hashtag

udf_sentiment_score

```
def sentiment_score(text):  
    blob = TextBlob(str(text))  
    score = blob.polarity  
    return score  
  
udf_sentiment_score = udf(lambda x:sentiment_score(x), FloatType())
```

Get Sentiment Score from TextBlob
Polarity ranged from -1 to 1

udf_sentiment

```
def sentiment(sentiment_score):  
    if sentiment_score > 0:  
        return "positive"  
    elif sentiment_score < 0:  
        return "negative"  
    else:  
        return "neutral"  
  
udf_sentiment = udf(lambda x:sentiment(x), StringType())
```

Categorization of Sentiment Score
(between "positive", "negative", and "neutral")

Transformed Data

username	date	tweet	sentiment_score	sentiment
GrabID	2021-10-11 23:09:02	If it on the poli...	0.0	neutral
GrabID	2021-10-11 22:59:46	Hi Sis :) for int...	0.25	positive
marsupilapila	2021-10-11 21:36:13	id want to ask fo...	0.0	neutral
alpyPutraa	2021-10-11 21:19:56	: Hello id I orde...	0.0	neutral
Naanganang	2021-10-11 21:13:43	id this min why s...	0.0	neutral
nanang_indra	2021-10-11 21:09:07	id hello sis rece...	0.0	neutral
cabangpurnamaa	2021-10-11 19:16:04	, I have to be fu...	0.0	neutral

Twitter

Transformed Data

userName	at	content	sentiment_score	sentiment
Ahmad Syukri	2021-11-28 12:09:21	Great ... Competi...	0.53333336	positive
sari gatot	2021-11-28 03:13:50	often throw the I...	0.0	neutral
Vera Puspa Indah	2021-11-28 01:23:53	Unfortunately , t...	-0.5	negative
Cimel	2021-11-28 01:06:59	good	0.7	positive
soetardjo shari	2021-11-28 01:06:23	Best application ...	0.65625	positive

Google Play Store



03

Load

Loading the transformed data into
data warehouse (AWS S3)

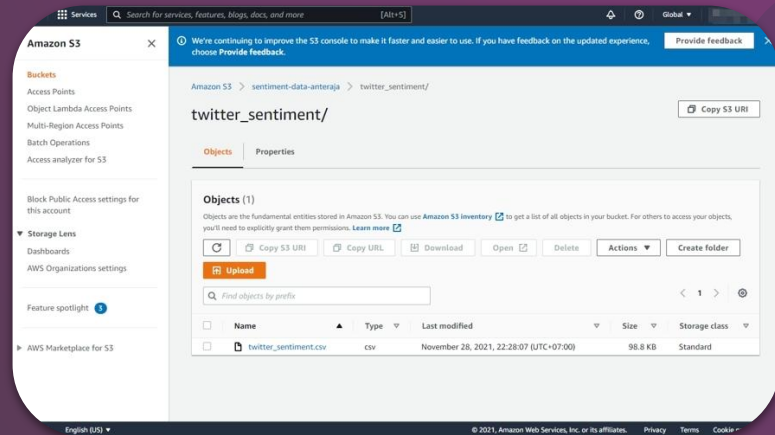
Load Twitter Data to AWS S3

```
# Load df_Twitter to S3
filename= "twitter_sentiment.csv"
bucketName = 'sentiment-data-anteraja'
folder_name = "twitter_sentiment/"

csv_buffer = StringIO()
pd_Twitter.to_csv(csv_buffer, index=False)

client = boto3.client("s3")

response = client.put_object(
    Body = csv_buffer.getvalue(),
    Bucket = bucketName,
    Key = folder_name+filename
)
```



twitter_sentiment.csv

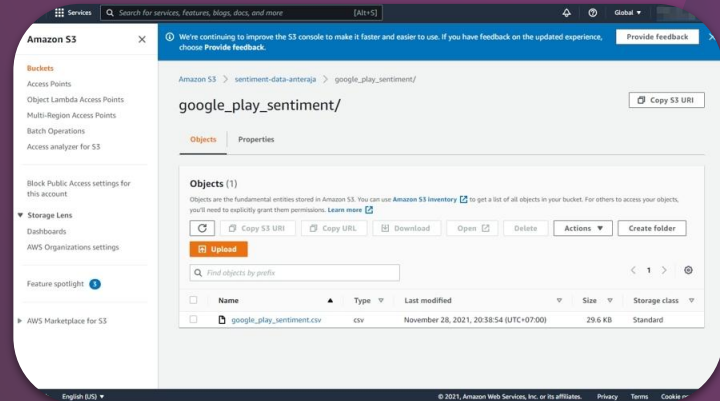
Load Google Play Data to AWS S3

```
# Load df_google_play to S3
filename= "google_play_sentiment.csv"
bucketName = 'sentiment-data-anteraja'
folder_name = "google_play_sentiment/"

csv_buffer = StringIO()
pd_google_play.to_csv(csv_buffer, index=False)

client = boto3.client("s3")

response = client.put_object(
    Body = csv_buffer.getvalue(),
    Bucket = bucketName,
    Key = folder_name+filename
)
```



google_play_sentiment.csv

AWS S3

sentiment-data-anteraja [Info](#)

Objects

Properties

Permissions

Metrics

Management

Access Points

Objects (2)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)



Copy S3 URI

Copy URL

Download

Open

Delete

Actions

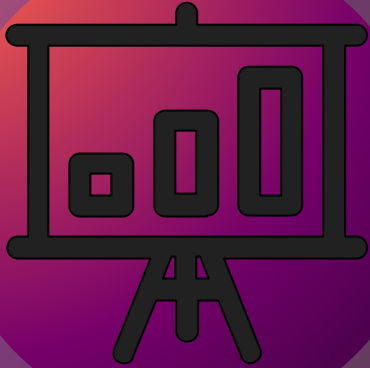
Create folder

Upload

Find objects by prefix

< 1 > ⚙

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	google_play_sentiment/	Folder	-	-	-
<input type="checkbox"/>	twitter_sentiment/	Folder	-	-	-

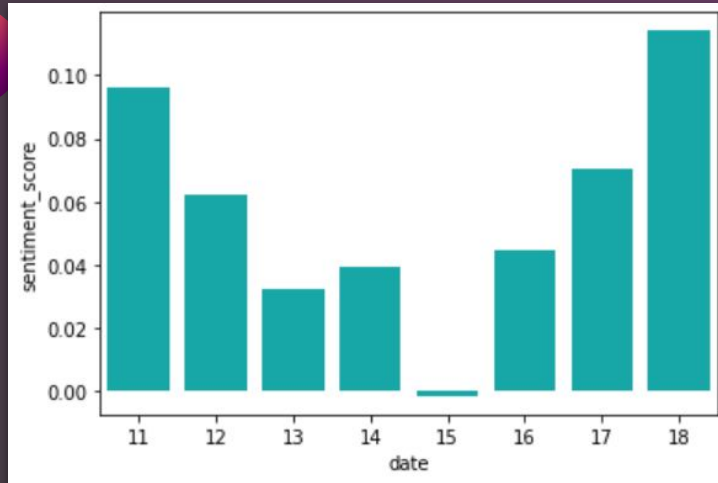


04

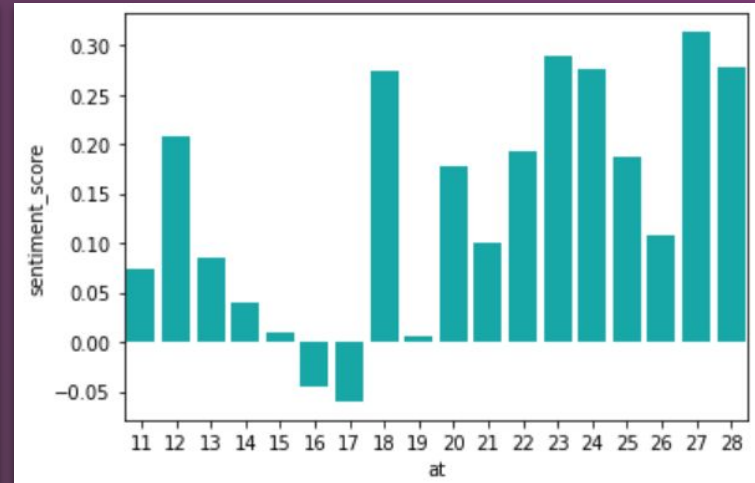
Visualization

Gathering insight from the twitter and
playstore review data

Visualization (Barplot)



Twitter



Google Play Store

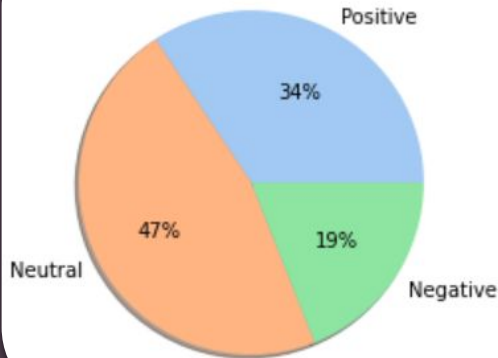
Visualization (Barplot)

Terdapat kesamaan penurunan sentimen pada data Twitter dan data Playstore, yaitu penurunan signifikan pada tanggal 15 November 2021. (Twitter = -0.001666; Playstore = 0.010236)

Hal ini diduga dikarenakan ekspektasi pelanggan mengenai waktu pengiriman selama 3-5 hari. Jika dihitung dari tanggal 11, maka 3-5 hari jatuh pada tanggal 14-16 November 2021. Dapat diduga penurunan tersebut diakibatkan oleh belum sampainya paket dikarenakan tingginya arus logistik setelah *event* 11.11.

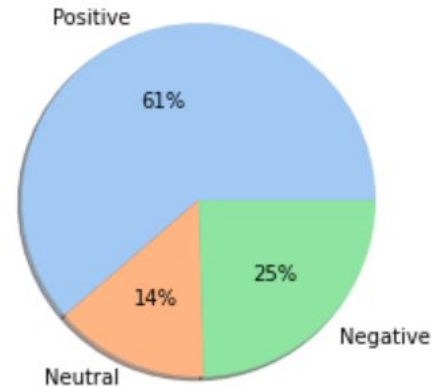
Visualization (Pie Chart)

Sentiment of tweets of Anteraja



Twitter

Sentiment of review of Anteraja on Google Play



Google Play Store

Visualization (Pie Chart)

Terlihat dari *pie chart* sebelumnya, sentimen negatif pada kedua sumber bukan merupakan sentimen yang dominan. Pada Twitter, sentimen yang dominan adalah *neutral* (sebanyak 47%), sedangkan pada Google Play Store adalah *positive* (sebanyak 61%). Hal ini dikarenakan sentimen negatif hanya terjadi pada beberapa hari setelah 11.11, yaitu pada saat pengiriman barang ke pelanggan.

Summary

The promotional strategy from e-commerce on November 11, 2021 led to an increase in logistics activities, one of which was Anteraja. Not a few customers use Anteraja's services for freight delivery. However, the delivery time is relatively longer than usual (up to 3 to 5 days) causing negative public sentiment as shown in user reviews.

Referensi

Aria, 2020. 11.11 dan Promo Rutin Tanggal Cantik, Strategi Shopee Dominasi Pasar.

<https://katadata.co.id/pingitaria/digital/5faa3b9049e9a/1111-dan-promo-rutin-tanggal-cantik-strategi-shopee-dominasi-pasar>

Asosiasi Logistik Indonesia, 2019. Transaksi Harbolnas 2019 Tercatat Rp9,1 Triliun. https://www.ali.web.id/web2/news_detail.php?id=1736

Thanks!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**