Customer Sentiment Analysis Based on Twitter

2021

Agenda

01 Sentiment Analysis

02 Methodology

03 Results

04 Conclusion

Sentiment Analysis

What is the sound of my product?

Sentiment Analysis

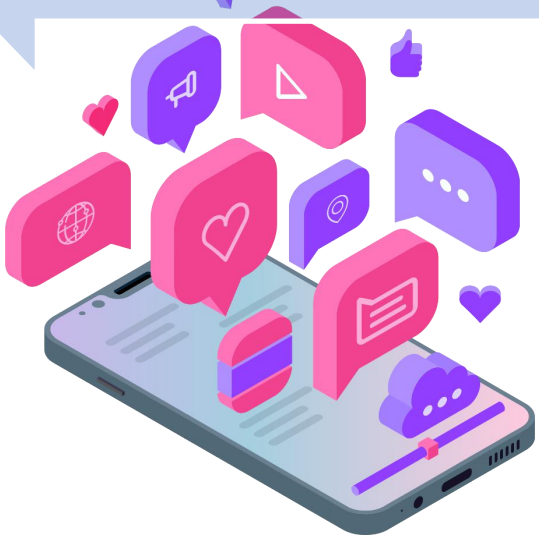
Capability of natural language processing, a kind of artificial intelligence. It could allow companies to search social media, the general web, and their backlog of customer support tickets for what their prospects and customers think about their brand and products.

54% of social browsers use social media to research products.

71% of consumers who have had a positive experience with a brand on social media are likely to recommend the brand to their friends and family.

49% of consumers depend on influencer recommendations on social media.

3.5 Billion
social media users worldwide 2020



How it is helpful in marketing?

Tracking Your Employees' Feedback

Improving Your Customer Support

Providing Better Product Analytics

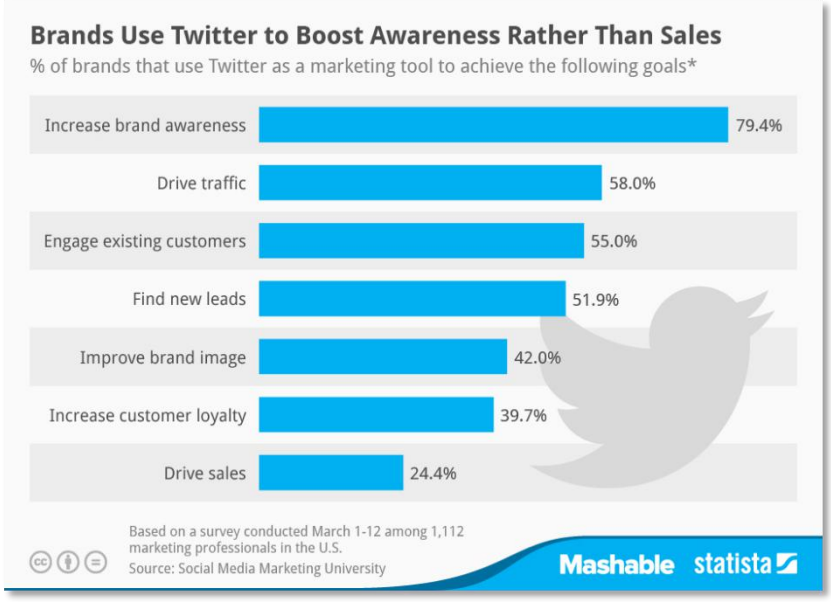
Brand Monitoring

Monitoring Market Research

How it is helpful in marketing?

Twitter listening allows you to investigate how and when your target users interact with your brand on this social network.

It opens up the opportunity to make precious conclusions on what types of content most engage your audience and at what time your customers are the most active online.



Methodology

How it works

SNA Workflow

01

In the first step, document are crawled from Social Media such as Twitter using any tool like Python, R, RapidMiner.

02

In this process, the given input document is processed for removing redundancies, inconsistencies, separate words, stemming and documents are prepared for next step.

03

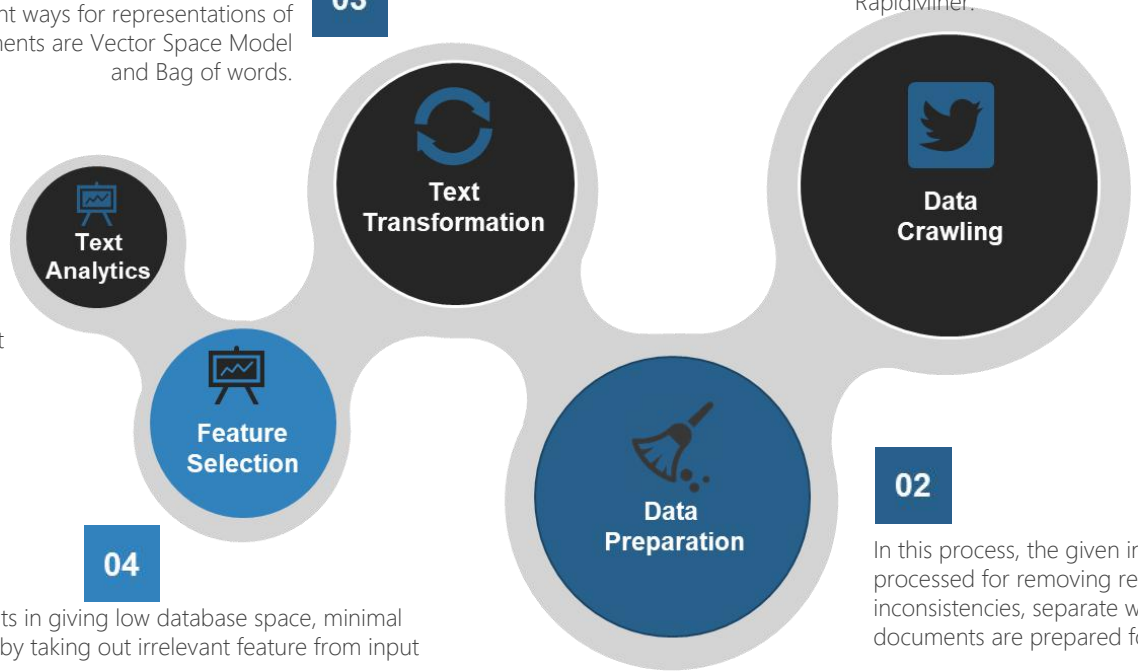
A text document is collection of words (feature) and their occurrences. There are two important ways for representations of such documents are Vector Space Model and Bag of words.

04

This method results in giving low database space, minimal search technique by taking out irrelevant feature from input document

05

In this stage the conventional data mining process combines with text mining process.



SNA Scenario for Kalbe

In this scenario, we will using sentiment analysis approach to learn about Hydrococo product customer sentiment to better understand audience and what they're saying on Twitter

Data

We will gather tweets about Hydrococo product

Request API

Request API to Twitter Developer for crawling, note that we can only gather data for last 7 days

Text Analytics

Processing tweets into meaningful information using sentiment analytics technique such as machine learning, lexical method, etc. then will apply for brand mapping

Dashboards

Creating dashboards to display meaningful information from text processing, information can be

- Most often words
- Sentiment scoring
- Top number of tweet location, etc

Data Crawling Stages

Data Source

GNIP or Datasift Provide pricing options to gives you more complete and entire archive of Twitter data

Crawler

Open source that able to gather data and powerful text mining tool for any language such as Bahasa

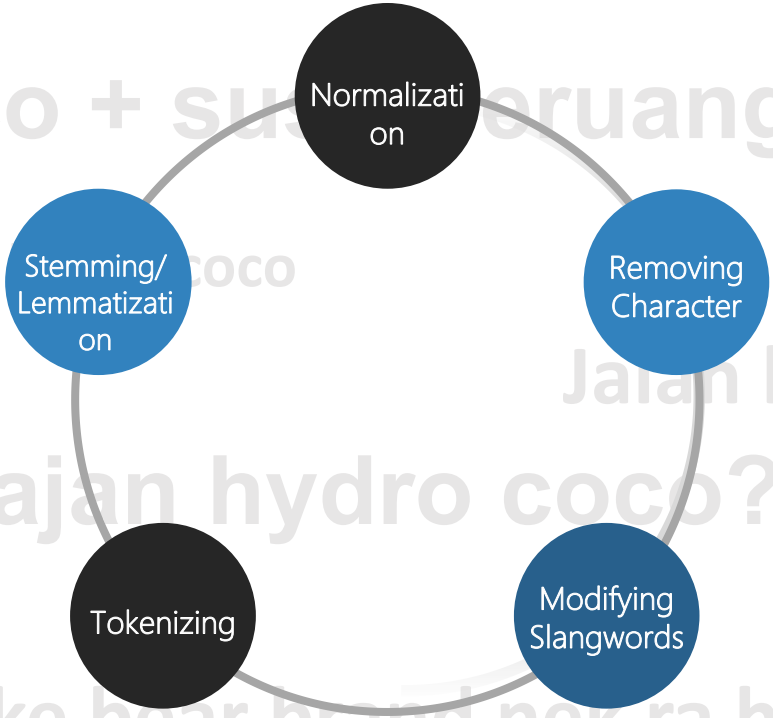
Storage/Speed Layer

Aims to minimize latency by providing real-time views into the most recent data and provide storage

Process

Using Text Processing Extension and Search Twitter Operator, able to gather data and apply text mining just point click. Provide free trial

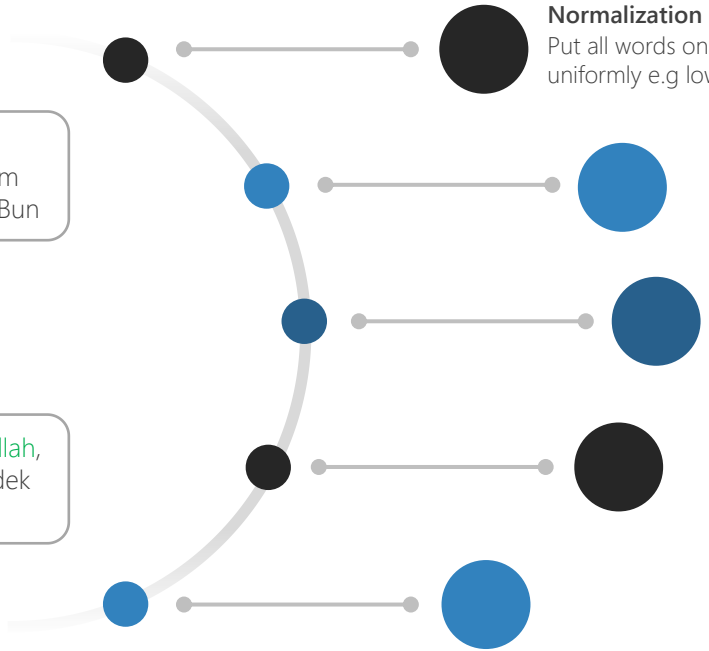
Data Preparation



Data Preparation

@tusssss_ @darjofess Alhamdulillah, semenjak minum Prenagen dedek jadi lebih aktif Bun ☺

@tusssss_ @darjofess alhamdulillah, semenjak minum prenagen dedek jadi lebih aktif bun ☺

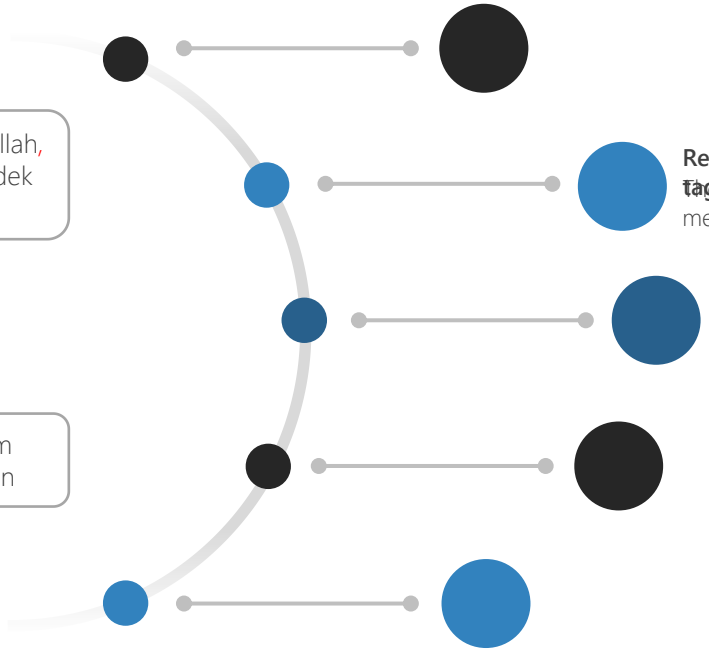


Normalization
Put all words on equal footing in order to allow processing uniformly e.g lowercase all text

Data Preparation

@tusssss_ @darjofess alhamdulillah, semenjak minum prenagen dedek jadi lebih aktif Bun ☺

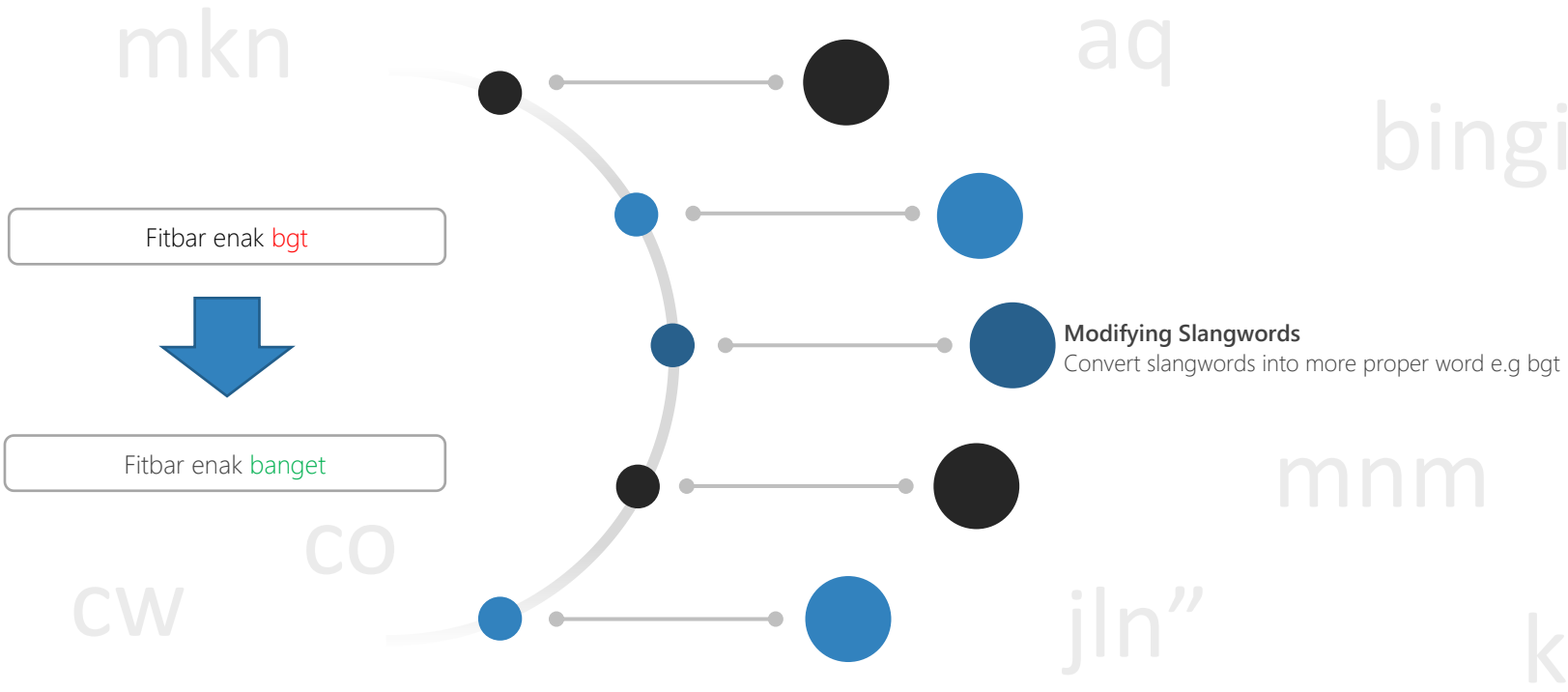
alhamdulillah semenjak minum prenagen dedek lebih aktif bun



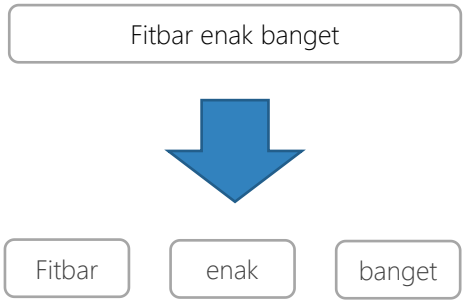
Removing Punctuation, Stopword, non –alphanumeric, Tags

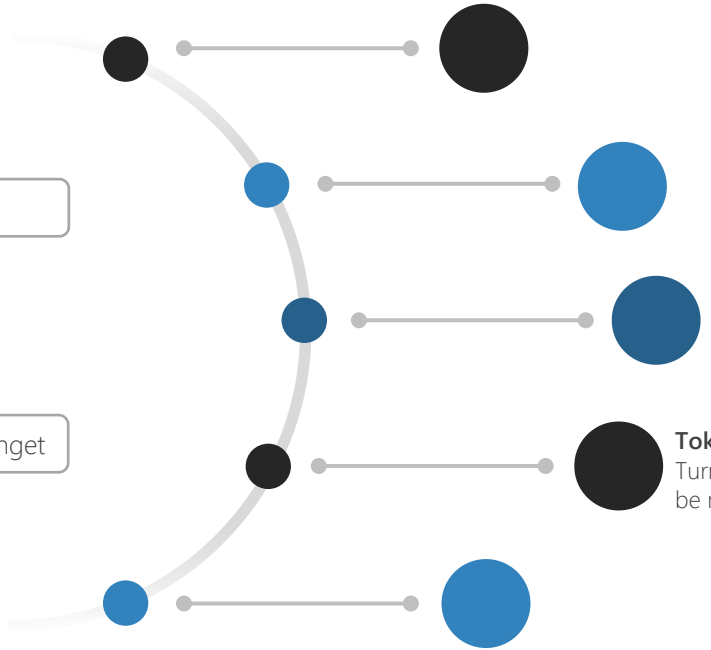
These are usually removed as they do not add much to the meaning of the text data e.g dari, adalah, @. ☺

Data Preparation



Data Preparation

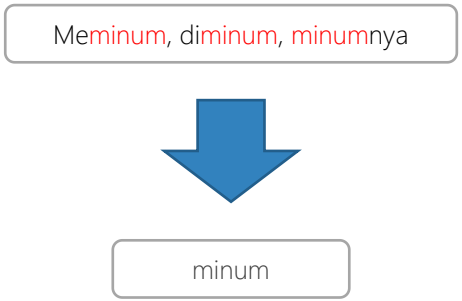


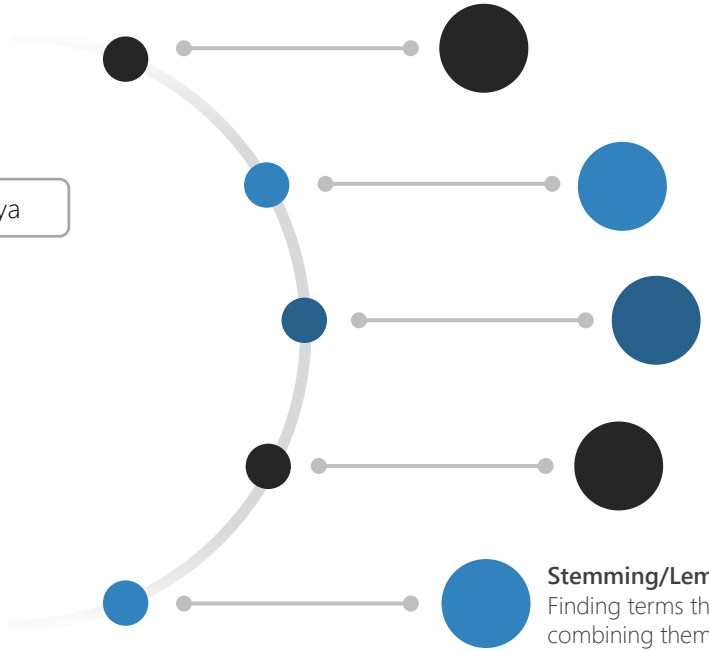


Tokenizing
Turn words in an input documents into attributes that can be mined, identified by separating characters as tokens

Data Preparation

Meminum, diminum, minumnya

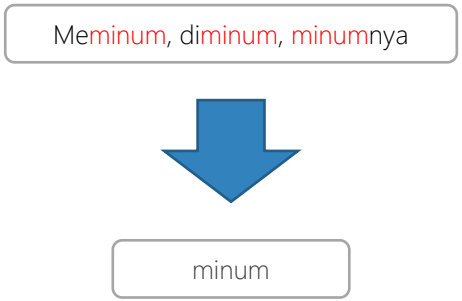


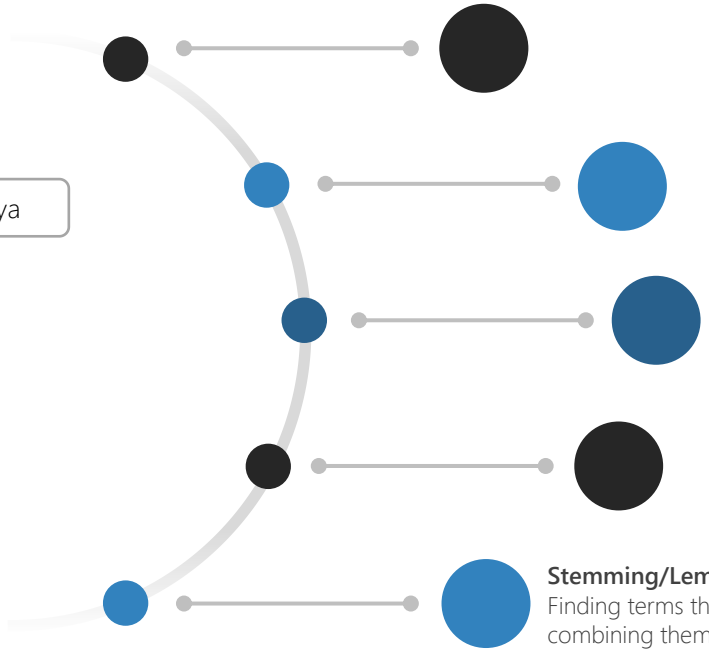


Stemming/Lemmatization
Finding terms that share a common word root and combining them because they mean essentially the same thing e.g minum, diminumkan, minuman > minum

Data Preparation

Meminum, diminum, minumnya





Stemming/Lemmatization
Finding terms that share a common word root and combining them because they mean essentially the same thing e.g minum, diminumkan, minuman > minum

Text Transformation

The core of text processing is turning the text data into quantitative data suitable for modelling. There are two ways represent of text data Bag of Words and Vector Space

Bag-of-words: the simplest form of text representation in numbers. It keeps count of the total occurrences of most frequently used words

Example
Review 1: This movie is very scary and long
Review 2: This movie is not scary and is slow
Review 3: This movie is spooky and good

	1 This	2 movie	3 is	4 very	5 scary	6 and	7 long	8 not	9 slow	10 spooky	11 good	Length of the review(in words)
Review 1	1	1	1	1	1	1	1	0	0	0	0	7
Review 2	1	1	2	0	0	1	1	0	1	0	0	8
Review 3	1	1	1	0	0	0	1	0	0	1	1	6

Text Transformation

Vector Space: Given the bag of words that you extracted from the document, you create a feature vector for the document, where each feature is a word (term) and the feature's value is a term weight. One of the best-known schemes is TF-IDF weighting.

TF-IDF : Gives higher weightage to important term and lesser weightage to unimportant term

$$TF(t) = \frac{Number\ of\ times\ term\ t\ appears\ in\ a\ document}{Total\ number\ of\ terms\ in\ the\ document}$$

$$IDF(t) = log_e(\frac{Total\ number\ of\ documents}{Number\ of\ documents\ with\ term\ t\ in\ it})$$

Term	Review 1	Review 2	Review 3	IDF	TF-IDF (Review 1)	TF-IDF (Review 2)	TF-IDF (Review 3)
This	1	1	1	0.00	0.000	0.000	0.000
movie	1	1	1	0.00	0.000	0.000	0.000
is	1	2	1	0.00	0.000	0.000	0.000
very	1	0	0	0.48	0.068	0.000	0.000
scary	1	1	0	0.18	0.025	0.022	0.000
and	1	1	1	0.00	0.000	0.000	0.000
long	1	0	0	0.48	0.068	0.000	0.000
not	0	1	0	0.48	0.000	0.060	0.000
slow	0	1	0	0.48	0.000	0.060	0.000
spooky	0	0	1	0.48	0.000	0.000	0.080
good	0	0	1	0.48	0.000	0.000	0.080

Text Analytics

There are broadly two categories of sentiment analysis which are Lexical Method and Machine learning method, in this scenario we apply lexical method first.

Lexical Methods :

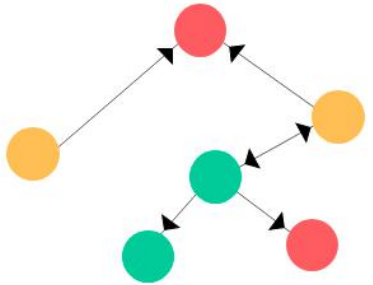
- These techniques employ dictionaries of words annotated with their semantic polarity and sentiment strength. This is then used to calculate a score for the polarity and/or sentiment of the document.
- We apply a lexicon-based approach in order to avoid the need to generate a labelled training set. Because the main disadvantage of machine learning models is their reliance on labelled data

	text	sentiment
0	@Penol86 Kurang. Hydro coco + susu beruang. Jos	1
1	Tokopedia. tp ak pake Shopee\r\nBank Bukopin. tp ak pake BCA\r\nSamsung. tp ak pake Oppo\r\nCoca...	3
2	@Zanubiasasaja Ombekke bear brand nek ra hydro coco nek ra guang kali wae	0
3	Jalan ke alfamart. Ngapain? Jajan hydro coco? 😂 https://t.co/t9hbQga4Us	0
4	@HydroCocoBITS 4 Bahan Pembuatan Tanaman Recycle :\r\n1. Botol bekas Hydro Coco Bits\r\n2. Tanam...	0

Text Analytics

➢ **Network Graph**

• Graph visualizes how subjects are interconnected with each other. Entities are displayed as nodes and the relationships between them are displayed with lines. This useful to summarize conversation in online social media services



➢ **WordCloud**

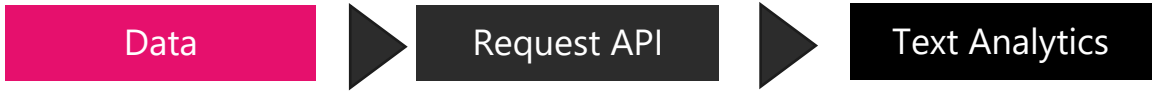
• A word cloud is an image composed entirely of words. The words vary in size and color, and the largest words suggest a higher frequency or usage. For example, brands can create a world cloud containing phrases they use in tweets

Sample

How it helpful

SNA Scenario Sample

In this scenario, we will be using some of social network analysis approach to learn about Hydro coco product customer sentiment to better understand audience and what they're saying on Twitter



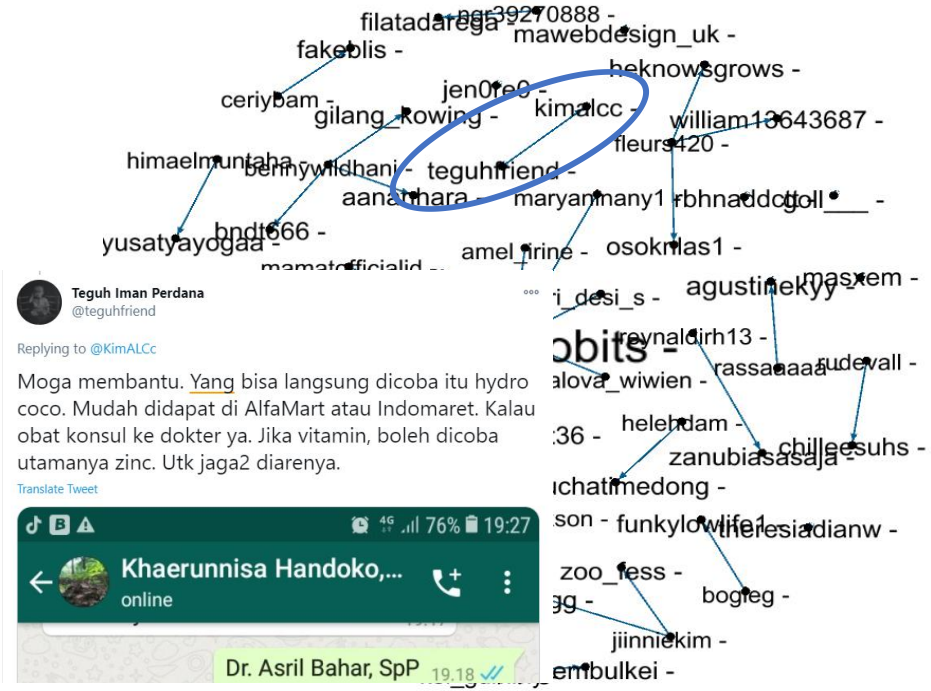
We will gather tweets about Hydrococo product with keyword #hydrococo, hydro coco, hydrococo

Request API to Twitter Developer for crawling, note that we can only gather data for last 7 days

Processing tweets into meaningful information using some method of text analytics which are Brand Mapping(Network Graph), WordCloud, Sentiment analytics

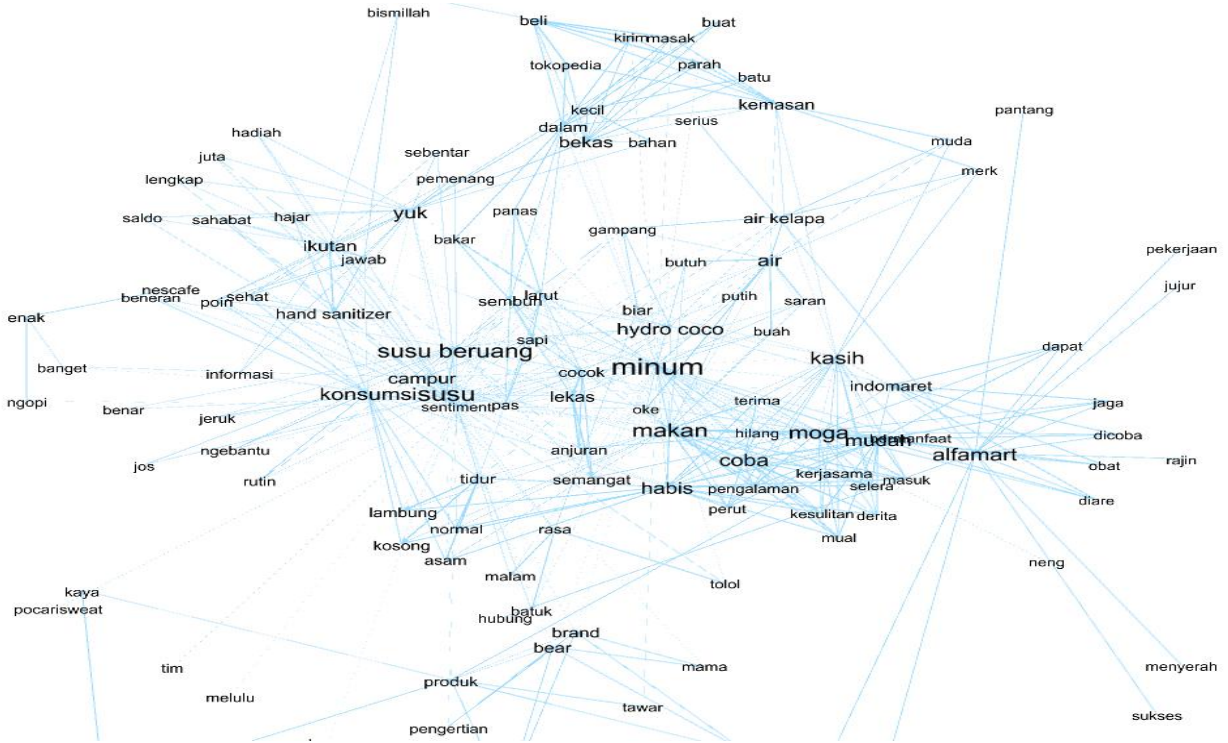
Support Campaign

Useful to see your customer by knowing what they are talking about, what kind of community talk about your brand and who they are

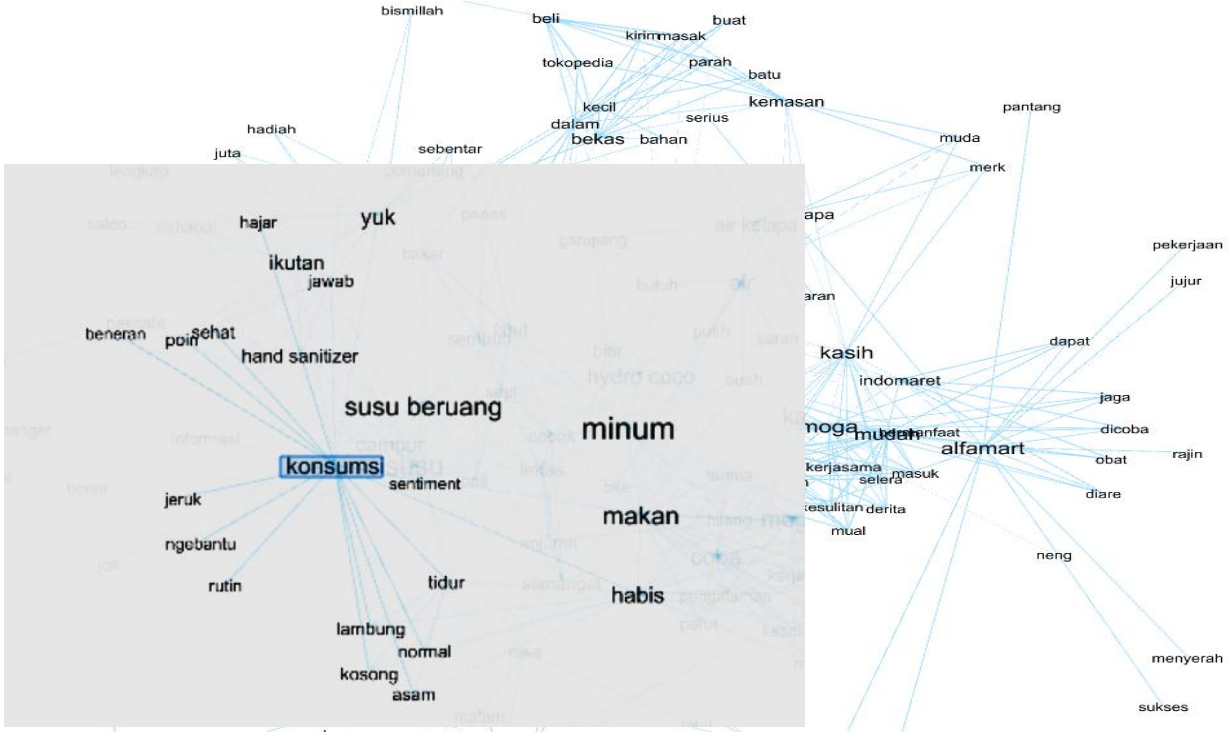


Association

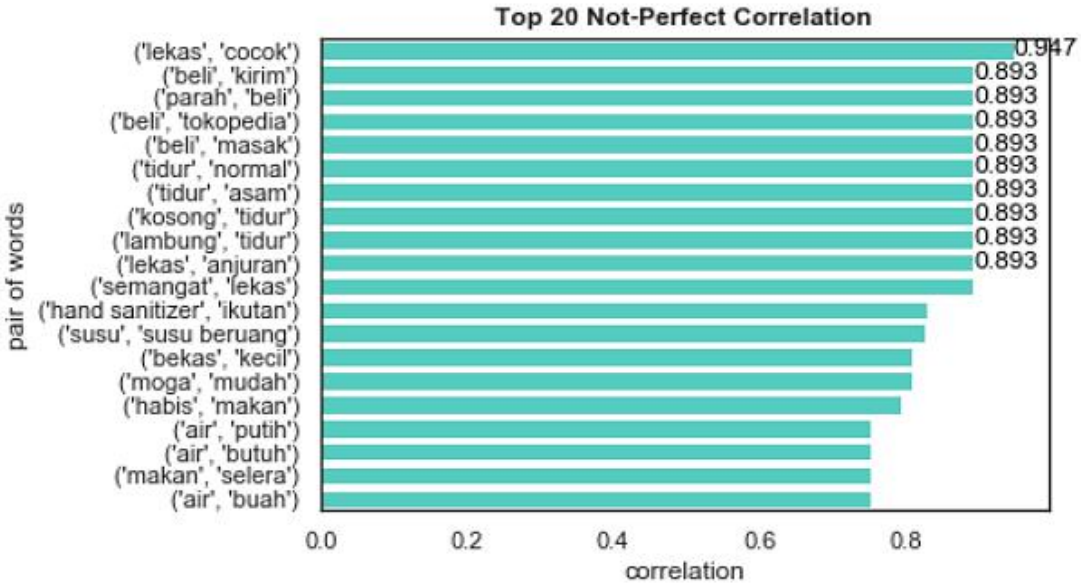
To know what words association to your brand, create associations in the minds of consumers which constitute brand image, and what other product associate with your brand



Association



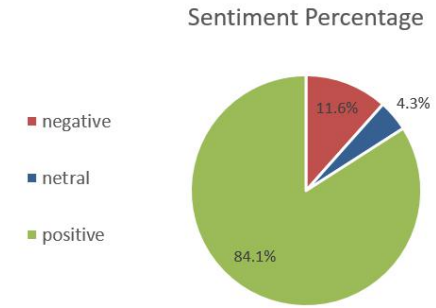
Catch Phrase

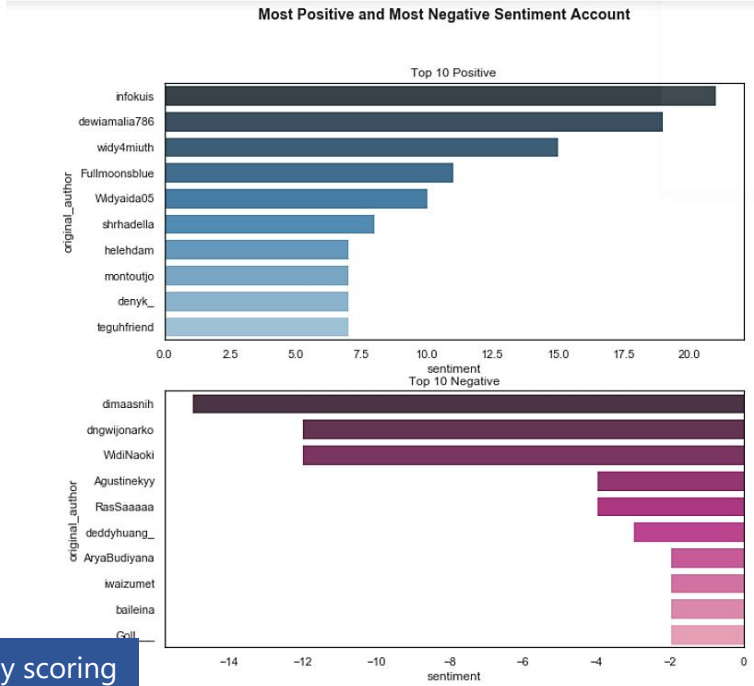


To see what most often word that occur in your product tweets, it is useful for social media marketing campaign manager to understand the product image

Campaign Success Rate

	text	sentiment
0	@Penol86 Kurang. Hydro coco + susu beruang. Jos	1
1	Tokopedia. tp ak pake Shopee\r\nBank Bukopin. tp ak pake BCA\r\nSamsung. tp ak pake Oppo\r\nCoca...	3
2	@Zanubiasasaja Ombekke bear brand nek ra hydro coco nek ra guang kali wae	0
3	Jalan ke alfamart. Ngapain? Jajan hydro coco? 😂 https://t.co/t9hbQga4Us	0
4	@HydroCocoBITS 4 Bahan Pembuatan Tanaman Recycle :\r\n1. Botol bekas Hydro Coco Bits\r\n2. Tanam...	0
5	@kochengfess Air kelapa nder, bisa hydro coco biar gampang nemunya jam segini atau susu beruang ...	2
6	@zelietmodista1 sudah pernah coba dicampur sama hydro coco belum Zell? https://t.co/NZsWZoPXKA	7
7	eruang beneran ngebantu\r\nHaja...	6
8	n3. Tanah\r\n4. Bebatuan kecil\r\n...	-2
9	a bisa yg kemasan merk hydro coco	-1





It is important to know what people feel about your product by scoring what they are talking about your product. Larger positive value means that the tweet is more positive, larger negative value means tweet is

Conclusion

Highlight Result

1. Network Graph user account shows the interaction between user accounts and it shows that @hydrococobits account has most interaction with other accounts, because it is sales account, while other account only has 2 or 3 interaction in average

2. Network Graph wordpairs shows the interaction between words, the larger the font size represent larger degree of a word,it means the word has most occurred with other words. The example is word “konsumsi” that occurred with recommendation keyword such as **yuk**, **ikutan** and a brand like ‘**susu beruang**’

3. Wordcloud shows several kind of recommendation keywords (e.g yuk, ikutan, hajar, lekas), some channels or brand (e. susu beruang, nescafe, pocarisweat, alfamart, tokopedia) , adjective (e.g enak, sehat, jos) and other keywords related to covid19 (e.g handsinitizer, masker), etc

4. Sentiment Scoring shows scoring value toward user tweet, it has mostly positive sentiment around 84%, while 11% is negative sentiment and 4% is neutral.

Conclusion

What Next ?

- Need more data sample to enhance the analysis, by gathering more data from other social medias like Facebook or Instagram or other product review website
- From wordcloud and network graph we see some brand or channel like ‘susu beruang’, ‘nescafe’, and ‘alfamart. If we have more data maybe we can map the relationship between this brand with product for competitive strategy analysis, etc.