

3. Influencers in social media

- Influencers have many followers setting trends in what people think or feel.
- Impacts sentiment aggregation – a core task of sentiment mining

Environment ministry pays social media influencers to spread word on climate change



4. Virality in social media

- Posts in social media go viral, sentiment. Causing eruptions in social media
- Bitcoin. A key factor in its rise end of publicity generated by viral
- Eg in cryptocurrency trading – of 2017 was the huge amount posts.

Bitcoin's soaring value was down to 'infected' buyers, economists say

Barclays analysts compare speculation in digital currency to spread of infectious disease



▲ Economists said Bitcoin's peak before Christmas was probably the ultimate price that could ever be achieved.
 Photograph: Dado Ruvic/Reuters

The rise of **bitcoin** has comparisons with the spread of an infectious disease, according to economists who argue the digital currency may have peaked in value as more consumers become immune to its appeal.

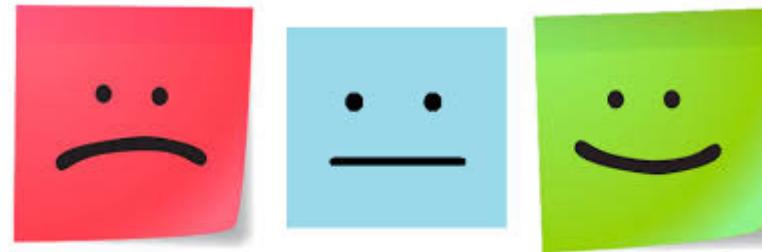
Summary

- Sentiment mining is a subset of text mining. It can be done at a document, sentence or even more granular aspect level.
- The issues regarding sentiment mining are non-trivial including most NLP-related tasks and other domain area issues.
- Its uses are extensive as long as there is text – in customer analytics, financial areas etc.
- New media has its own nuances in impacting sentiment – influencers, spam, virality and fake news.

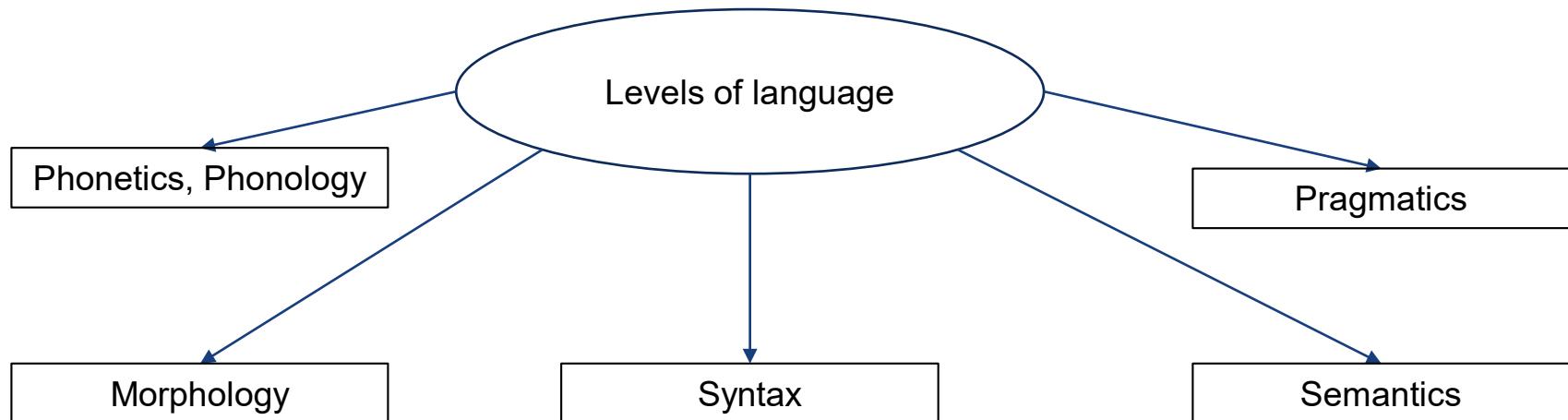
References (Introduction)

- Liu Bing, Sentiment Analysis and Opinion Mining, Morgan & Claypool, May 2012, available for free download at
<http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.html>
- Recommended Readings
- Bo Pang & Lilian Lee, Opinion Mining and Sentiment Analysis, in Foundations and Trends in Information Retrieval Vol 2 No 1-2 (2008). Prepublication version available for free download at
<http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf>
- Liu Bing, Sentiment Analysis and Subjectivity, in NLP Handbook, 2nd Ed., eds: N.Indurkhya & F.J.Damerau, 2010. Draft available for download at <http://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf>
- Speech and Language Processing Aug 2020
<https://web.stanford.edu/~jurafsky/slp3/>
- These references mainly describe the nuances that are particular to sentiment mining and are relatively 'outdated' (before ~2013). Still there are important principles to learn from it.

Types of Sentiment Mining



Introduction to Language Structure

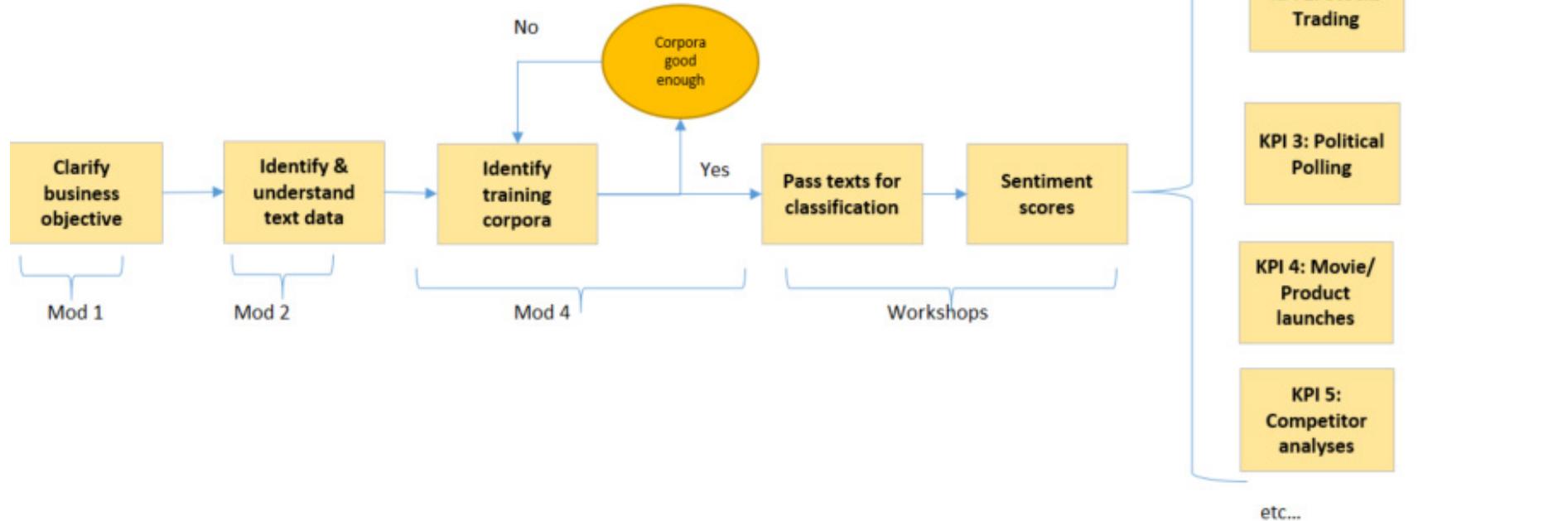


| Name of field (Top down approach) | Objective of study | Type of NLP task |
|-----------------------------------|---|--|
| Phonetics | All human sounds | Speech recognition |
| Phonology | Classified sounds | |
| Morphology | Words, forms | Lexical analysis, Lemmatisation, stemming, POS tagging |
| Syntax | Syntactic structure: Sentences, clauses | Parsing, Chunking |
| Semantics | Representation structure: Meaning | Entity extraction, word sense disambiguation, sentiment analysis |
| Pragmatics | Communicative intentions | Open Information extraction |

Overview of Sentiment Analysis System

Business objectives met? What actionable insights?

Overview of Sentiment Analysis System



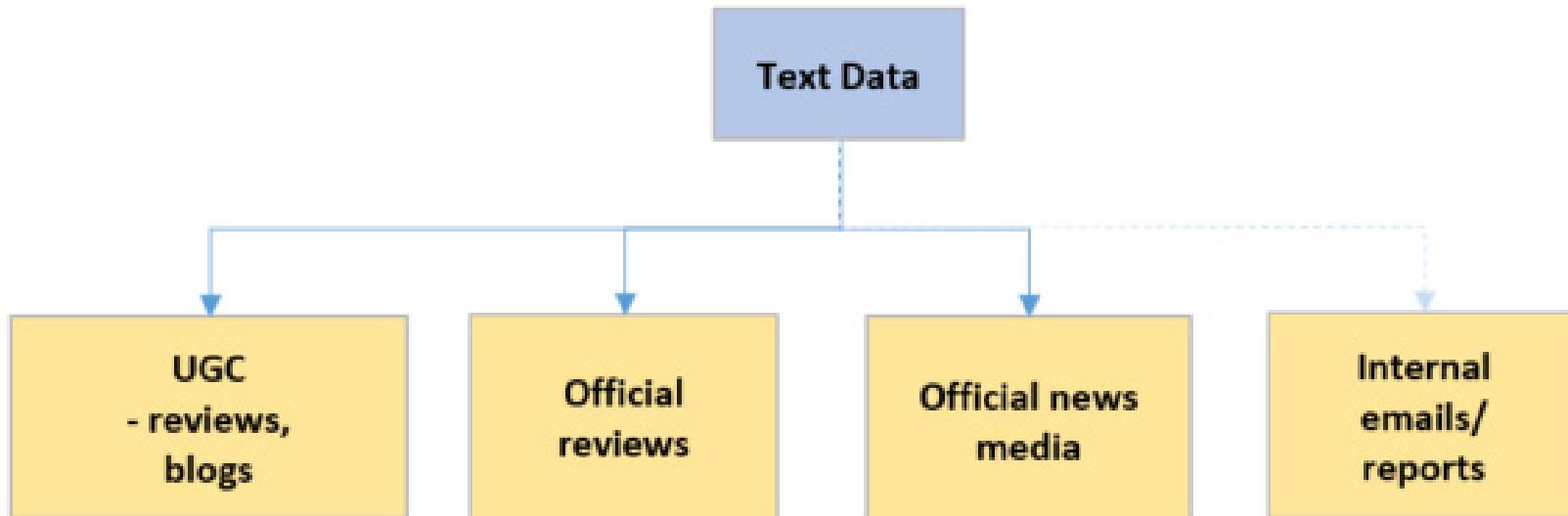
Sentiment analysis system

- In any business analytics system, the **business objectives** need to be defined. This needs to translate to a *visible ROI*. Example of business objectives are:
- Increased customer satisfaction resulting in less churn, better customer prospecting
- Increase in trading profits from sentiment analysis trading strategies.

Can you name other business objectives??

Sentiment analysis system

- Next step is identification of **data source** and **type of text data**. The pre-processing pipeline will be very much dependent on type of text data.
- Ex: RSS feeds, web scraping.
- Webpage titles provide a ‘gist’ of the content.
- Social media from media like WSJ, Reuters are usually well-written.



Sentiment analysis system

- Before training the sentiment algorithm classifiers, the **training corpora is KEY.**
- Experience shows it is more important than the algorithms itself. Garbage in garbage out is very true.
- Training corpora (body) needs to be as similar as possible to the test set. Needs to continually update this training corpora.
- This will be very much aligned with the business goal.

Sentiment analysis system

- **Return on Investment (ROI):** There must be *actionable insights* from the generation of sentiment scores
- These insights must be well communicated to stakeholders via sentiment visualization (ex: dashboards).
- Ex of actionable insights Identify under-valued stocks to buy?
- Identify customers feeling unhappy about services/ products?

NLP tasks in sentiment analysis (Quick revision)

- Sentiment analysis involves heavy NLP tasks. *Pre-steps are essential in the extraction for **training features**.*
- The following are the pre-NLP tasks for sentiment analytics (covered in text mining course)
 1. Tokenization (**Can be dependent on data and domain**)
 2. Stemming / Lemmatization (**Choose sentiment aware version**)
 3. Stop words (**Depends upon feature extraction method**)
 4. Part-of-speech (**Depends on Corpus**)
 5. Entity recognition (**Depends on Corpus**)
 6. Other NLP tasks – co-reference resolution, word sense disambiguation etc..
 7. Parsing (shallow, deep, dependency) – (**Important task**)
 8. Entity linking (**Very useful for information retrieval systems & Knowledge graph related task**)

End Thoughts

- A sentiment analysis system as with any business analytics projects needs to define the business objectives and have tangible ROI.
- Sentiment analysis involves many NLP tasks. There are subtle ways how these NLP tasks can impact the sentiment scores/ analysis.

Type of Sentiment Analysis

Fine-grained Sentiment Analysis

- If polarity precision is important to your business
- **Positive, Neutral, Negative**
- We can use **5-start** rating to derive labels if data is not labeled

Emotion Detection

- Aims to detect emotions, like happiness, frustration, anger & sadness etc.
- Lexicon based or Machine learning based
- Downside of Lexicon based systems: It is less context aware
- Ex: **Your customer support is killing me , you are killing it**

Aspect-based Sentiment Analysis

- If knowing which particular feature or aspect people are mentioning and its polarity is important to you business
- "The battery life of this camera is too short" - an aspect-based classifier would be able to determine that the sentence expresses a **negative** opinion about the feature **battery life**

Tools Available

Open source

(Make sure to know about the support and community around any tool before deployment. Also read the terms and conditions if you're using them in deployment (for language models etc.)

Proprietary

(If available in your organization, then don't re-invent the wheel. First try them & improvise them as per your need. They can be production ready and can be easily complemented with open source)

| Name | Features | Remarks |
|---|--|--|
| NLTK (very well known and vast resources) | More than 50 corpora and lexical resources, sentiment analysis, NER, POS, tokenizer and topic segmentation | Well documentation, community, large no of language support. |
| Spacy (Enterprise grade and very fast) | Pre build models for NER, POS and classification etc. | Fastest one and community is expanding very fast. It's resource intensive. |

| Name | Features | Remarks |
|-----------|--|---|
| Microsoft | POS, tokenizer, sentiment, spellings, language, topic modelling, multi lingual support | Industry ready and very easy to integrate with Microsoft services |
| Google | Auto-ml for NLP, POS, NER, Graphs support, Sentiment analysis, language detection, Word2Vec original best in class | One of the best in class and used by many. Available demos |

Tools Available

Open source

(Make sure to know about the support and community around any tool before deployment. Also read the terms and conditions if you're using them in deployment (for language models etc.)

Proprietary

(If available in your organization, then don't re-invent the wheel. First try them & improvise them as per your need. They can be production ready and can be easily complemented with open source)

| Name | Features | Remarks |
|---|---|---|
| TextBlob (easiest one and rapid prototyping) | Sentiment analyser (rule based & ML based), Google translate API, POS, tokenizer, spelling correction etc.. | Light weight and accessible. Utilizes NLTK pre-build models and pattern web mining module |
| Gensim | Topic and vector space modelling, multipliable classification & dimension reduction | Open source community and commercial support available. scalable and speedy |

| Name | Features | Remarks |
|-------------|--|---|
| AWS | Translation, sentiment, POS, NER, Topic modelling and Ontologies support | Promising and well backed by best in class infrastructure |
| IBM Watson | Sentiment, classification, ontologies, topic modelling and NER | Very well learned in some domains |
| Rakuten-nlp | word segmentor + PoS Tagger, tokenizer | Specifically for Japanese & Mandarin |

Tools Available

Open source

(Make sure to know about the support and community around any tool before deployment. Also read the terms and conditions if you're using them in deployment (for language models etc.)

Proprietary

(If available in your organization, then don't reinvent the wheel. First try them & improvise them as per your need. They can be production ready and can be easily complemented with open source)

| Name | Features | Remarks |
|---|--|--|
| Stanza (Stanford core NLP python implementation) | Tokenizer, sentiment analyser, POS, Parser etc.. | Another strong contender, many languages supported & have live demos. Famous in production systems |

| Name | Features | Remarks |
|----------|--|---|
| BaiduNLP | lexical analysis, Dependency parsing, Word vector representation, sentiment analysis | Better for language specific translation. |

Other useful communities and some demo links:

<http://textanalysisonline.com/>

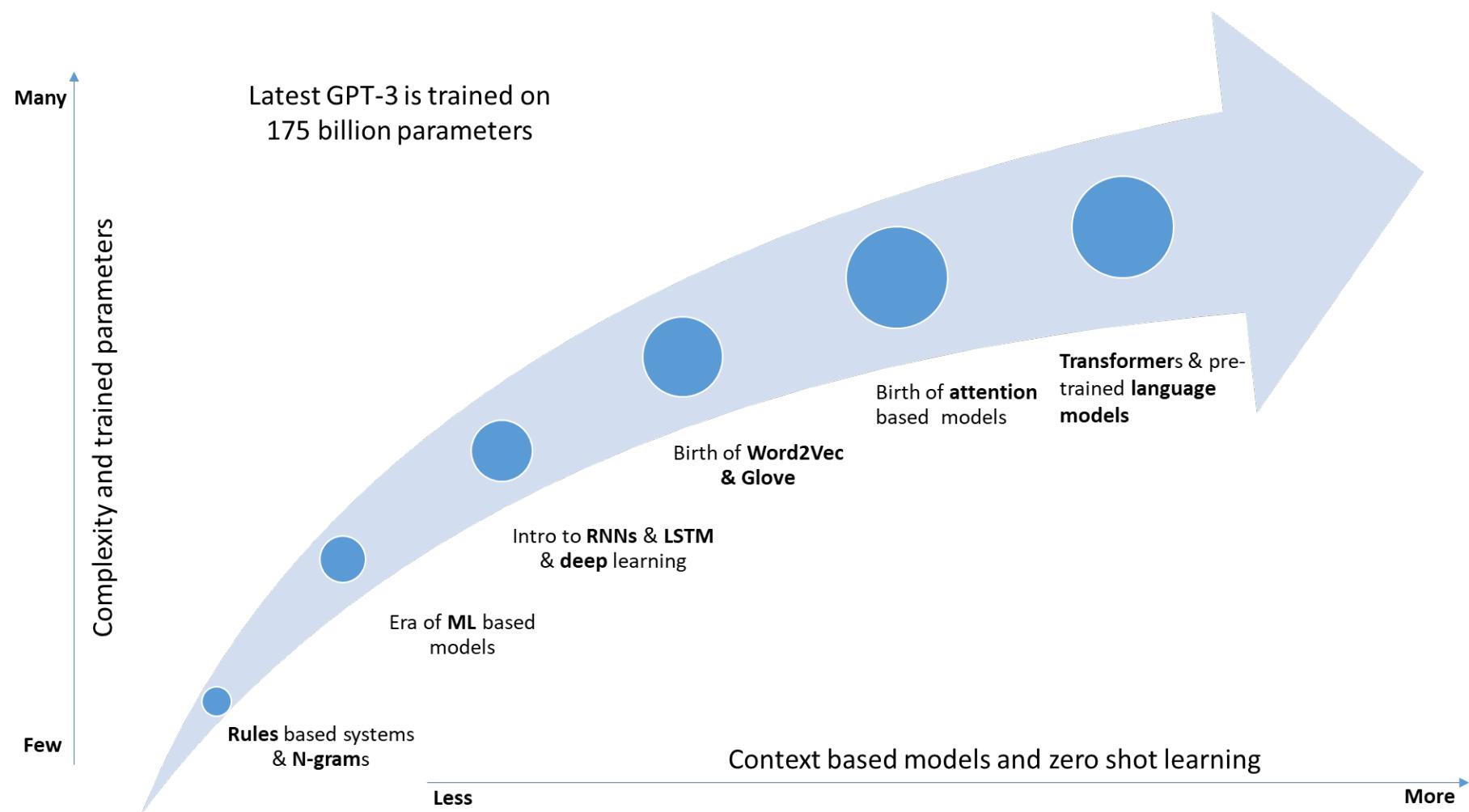
<https://api-inference.huggingface.co/docs/>

<https://demo.allennlp.org/sentiment-analysis>

<http://stanza.run/>

Other Libs: Textacy (used before and after Spacy)
 Pattern, Polyglot, Affin, MontyLingua

Evolution of NLP Models



Real Life Example

- <https://duckduckgo.com/?q=gramerly+email+tone&atb=v247-1&iax=images&ia=images&iai=https%3A%2F%2Fi.insider.com%2F5fb5450083df950018ac20d4%3Fwidth%3D600%26format%3Djpeg%26auto%3Dwebp>
- <https://quillbot.com/>

Why build a sentiment engine?

- There must be *actionable insights* from the generation of sentiment scores
- These insights must be well communicated to stakeholders via sentiment visualization (example dashboards).
- Example of actionable insights
 - Identify under-valued stocks to buy?
 - Identify customers feeling unhappy about services/products?

Sentiment summarisation

- Sentiment summarisation:
 - Aggregates sentiment from different texts over time to communicate insights
- Recall the quadtuple(o, t, t, s)from sentiment definition and quintuple (o, a, t, t, s)for aspect-based sentiment analysis?
- Summarization is done on three main levels:
 - the overall sentiment/ opinion across the writers (o)
 - the sentiment scores across time (t)
 - The aspects(a) or entity targets talked about(t)

Sentiment aggregation (opinion holders)

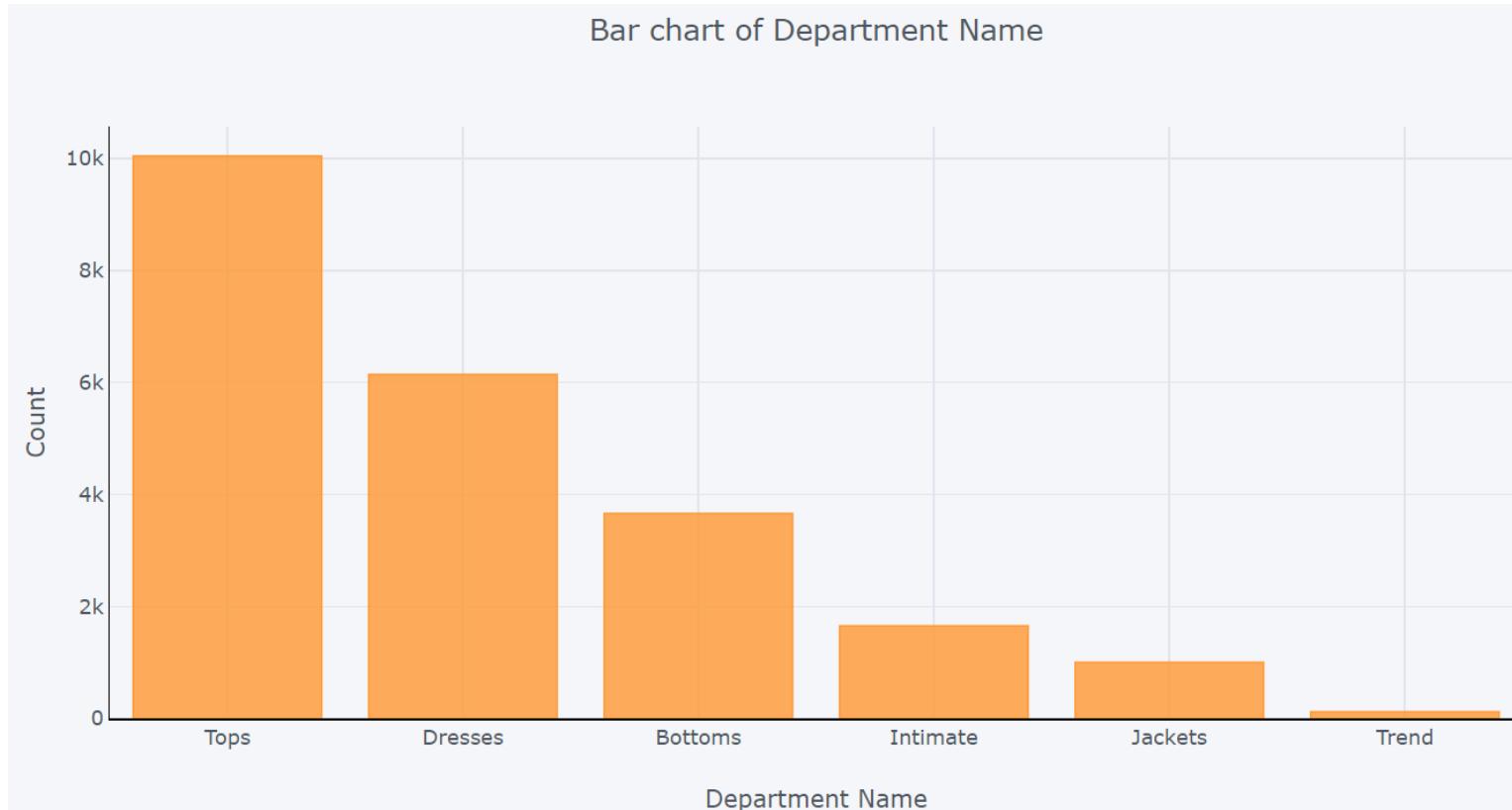
- Aggregating over opinion holders, differing weights:
 - Which opinion holders is most important in influencing opinions?
 - What are the factors to consider?
- Consumer marketing
 - Influencers (likely to cause virility)
 - No of tweets/ re-tweets (from meta-data)
- Financial news
 - Types of news: regular or extraordinary reporting
 - News source: more reputed sources WSJ or social media?

Sentiment aggregation (over time)

- Aggregating over time
 - This needs to be considered with the trading strategies and sentiment decay.
 - Trading horizon:
 - Often sentiment is cited for intra-day trading. In this case, should aggregate over shorter time-frame?
 - Also, need to tie in with the advertising campaign. Over days or weeks?

Sentiment Aggregation

- Dimensions of how you want the sentiment to be aggregated by:
 - Time, organisational units, aspects, products etc.
 - Metrics can be both sentiment and no. of comments



Categories and sentiments

