Ollscoil na Gaillimhe
University of Galway

Introduction to NLP

Social Media Analysis

Dr. Paul Buitelaar
Data Science Institute, University of Galway

Includes updates by Omnia Zayed

Learning Objectives

- Understand main objectives and challenges in social media analysis
- Understand challenges in and approaches to sentiment analysis
- Understand main ideas of other social media analysis tasks: emotion analysis, offensive content identification, misinformation detection

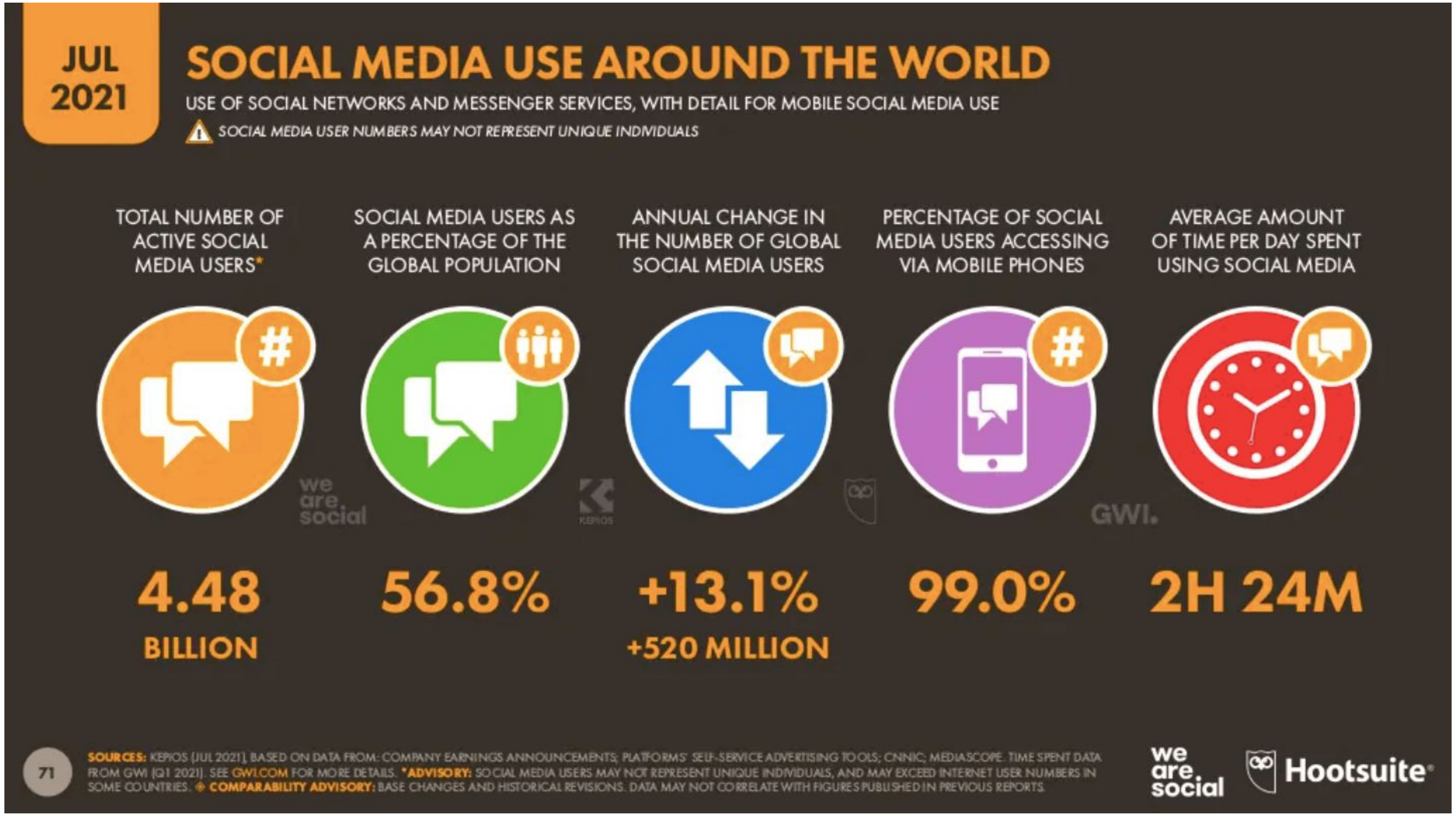
Social Media Analysis

This lecture is about opinion mining from social media

- **social media data**: type of data
- **opinion mining**: type of analysis

Images source: https://www.dreamstime.com

Importance of Social Media Data



Opinion Mining

Remember from Lecture 1

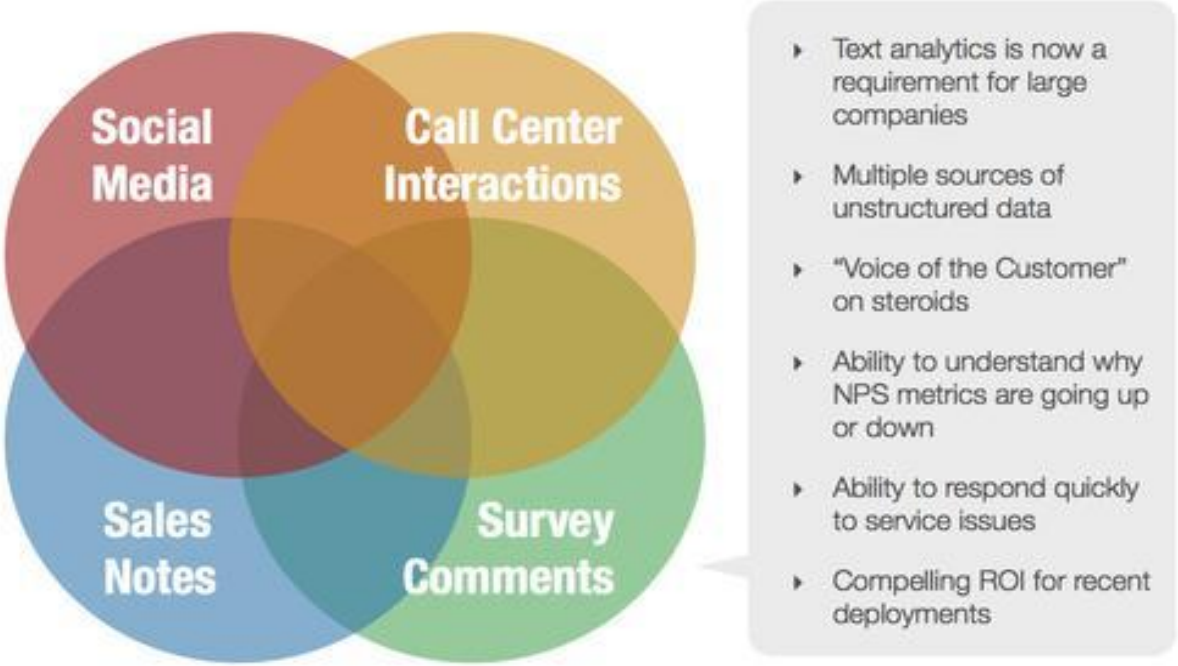


Image source https://www.linkedin.com/pulse/voice-customer-tells-tale-part-2-chris-marshall/ ; https://www.gracenote.com/

Opinion Mining – Some Definitions

Opinion mining is concerned with **analyzing opinionated text** in terms of sentiments, emotions, suggestions, arguments, …

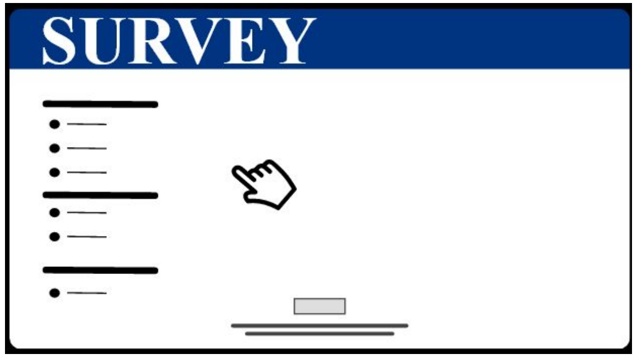
Please note: ‘opinion mining’, ‘sentiment analysis’, ‘subjectivity classification’, ‘emotion analysis’ are often **used interchangeably**

Alternatively, **opinion mining is used as umbrella term** for these different tasks

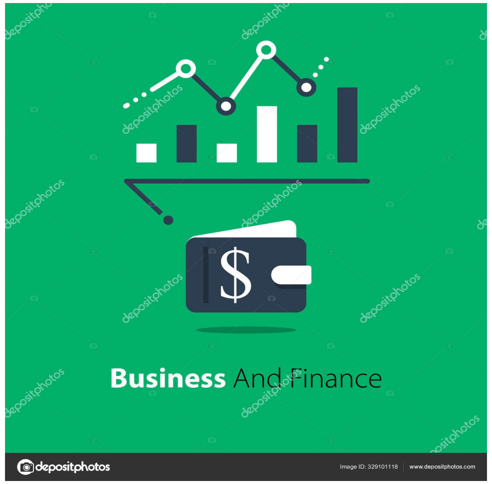
Opinionated Text - Some Examples

Reviews

Debate



Surveys



Analysis

Image sources: https://en.wikipedia.org/wiki/File:Debate_Logo.svg; https://blog.hubspot.com/service/customer-review-sites; https://depositphotos.com/329101118/free-stock-illustration-financial-analysis-business-performance-report.html; https://commons.wikimedia.org

Ollscoil na Gaillimhe
University of Galway

Buzz Groups

Can you think of any applications of opinion mining (e.g. through social media analysis)?

Sentiment Analysis on Products/Services

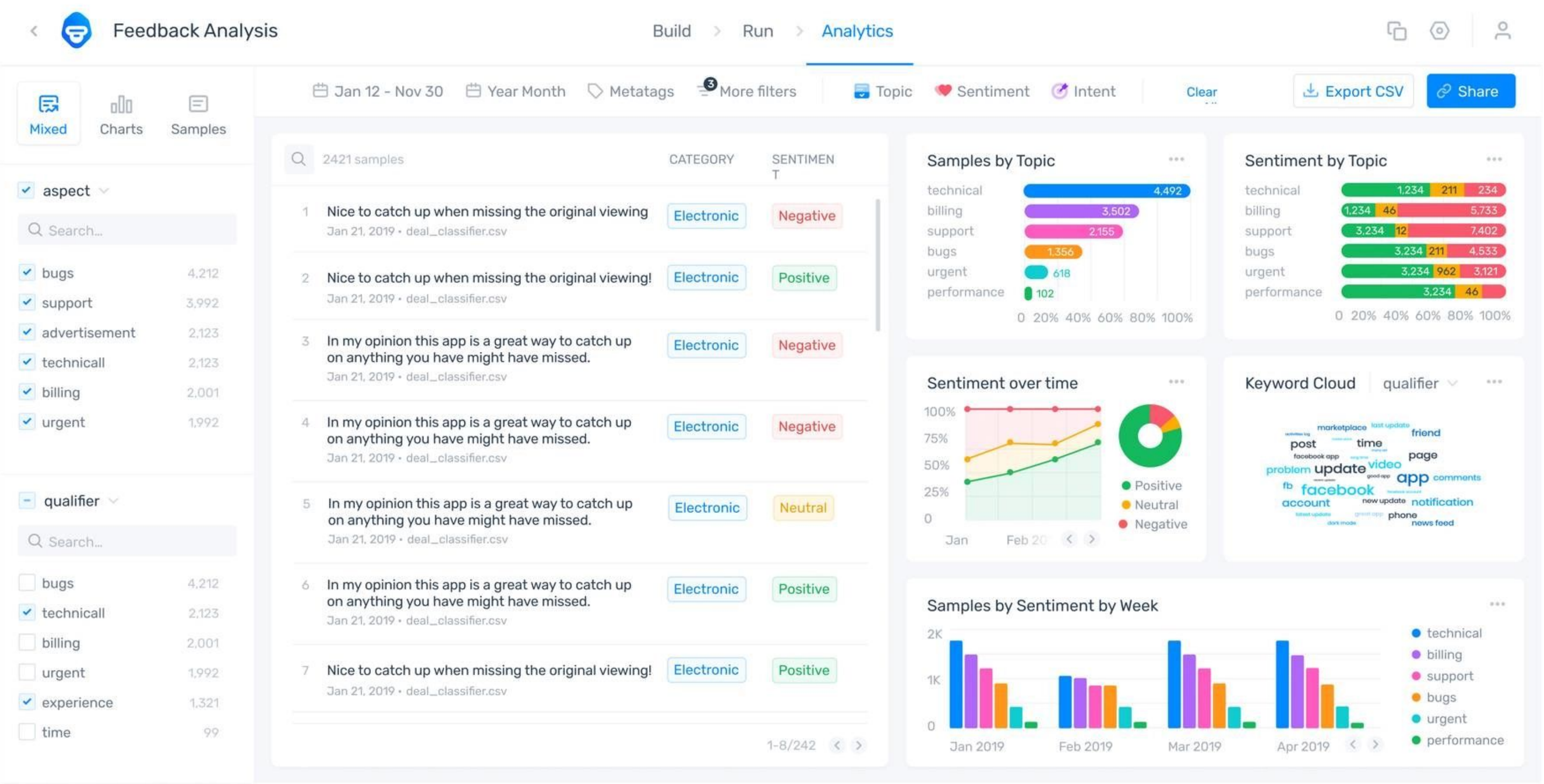
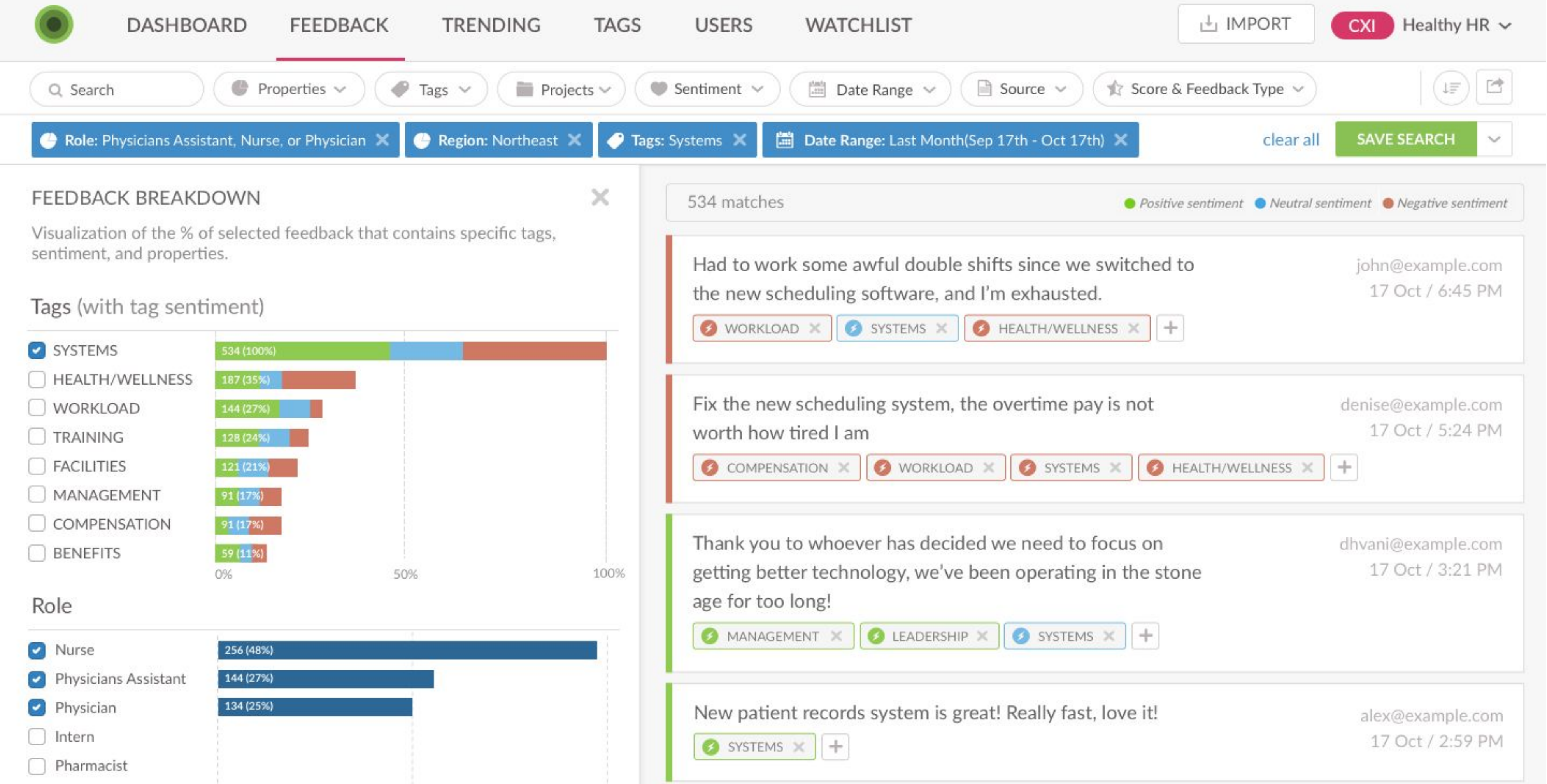
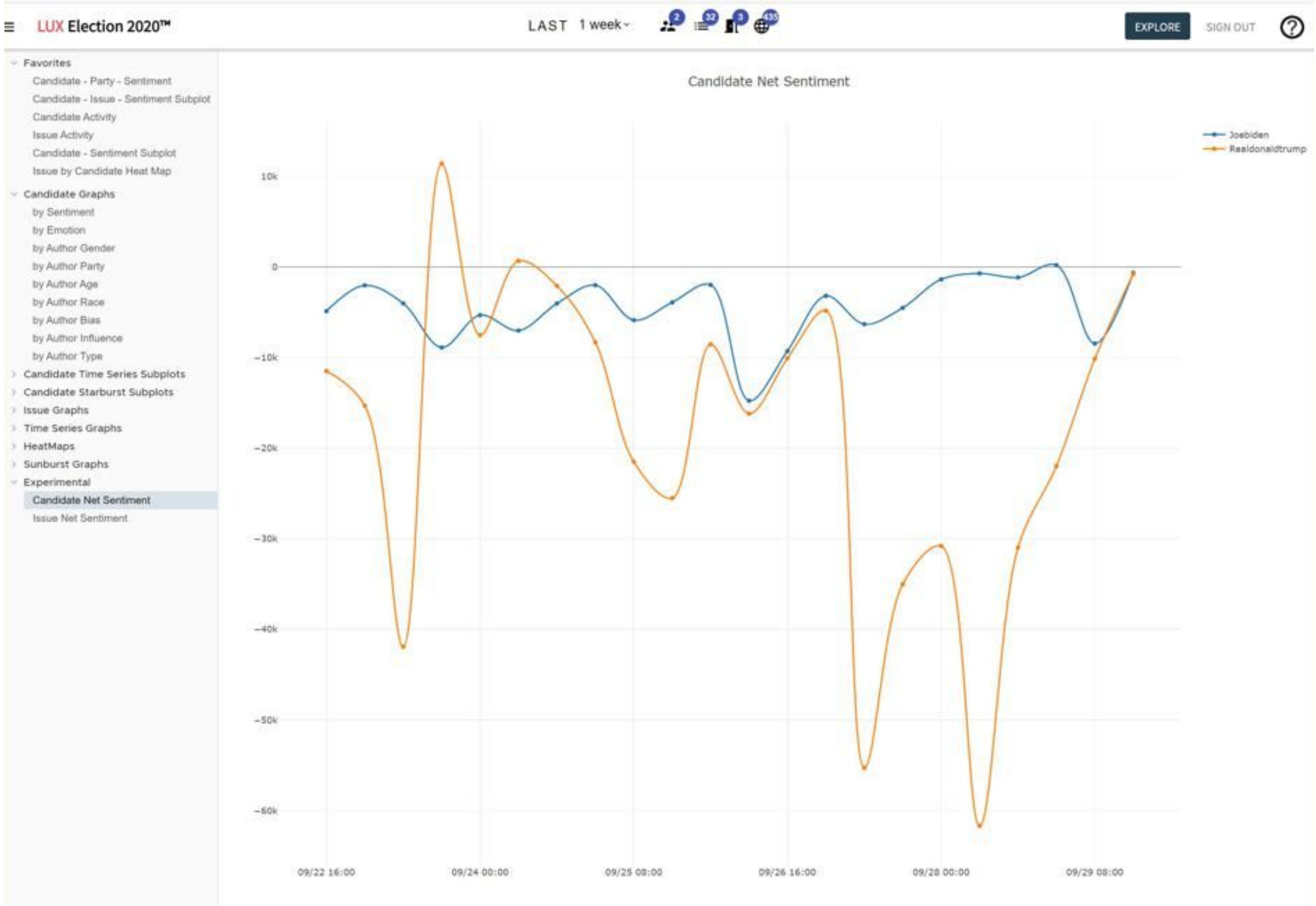


Image source; https://monkeylearn.com/blog/sentiment-analysis-of-twitter/

Sentiment Analysis in the Enterprise

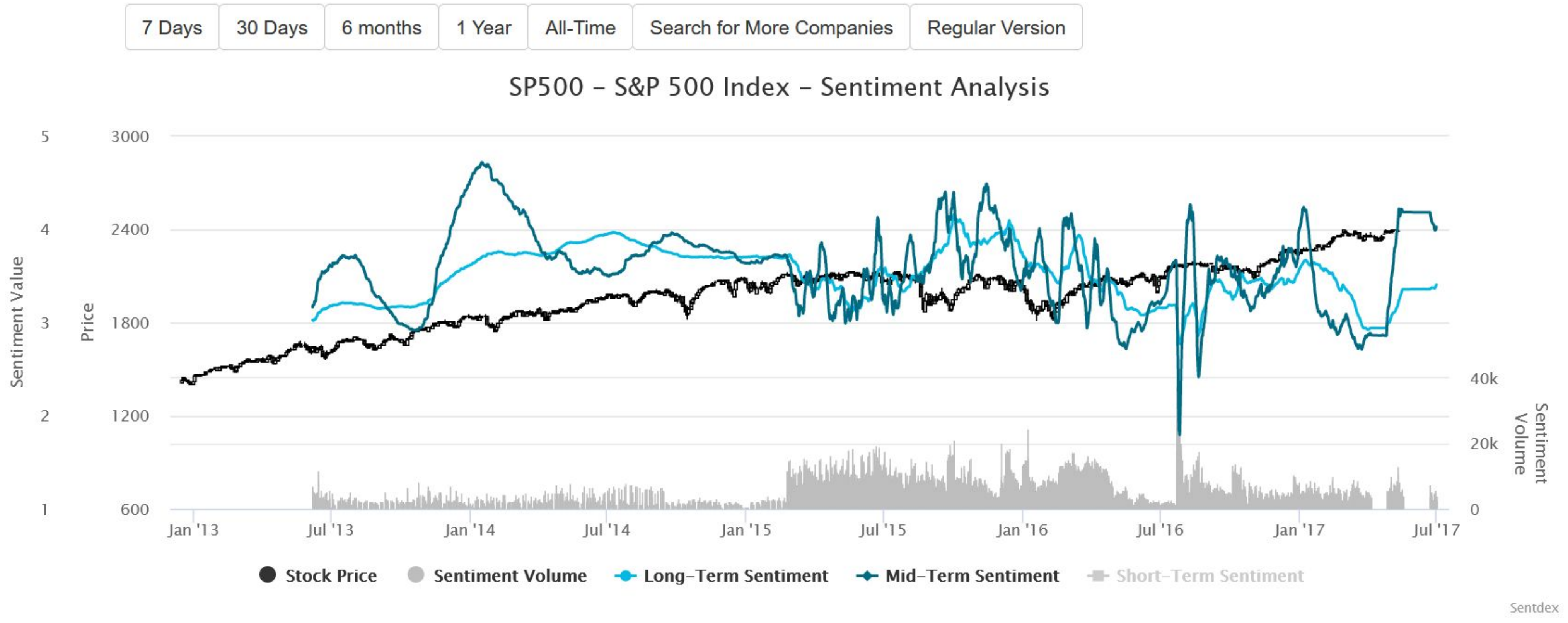


Sentiment Analysis in Politics



https://www.forbes.com/sites/waynerash/2020/09/29/new-big-data-sentiment-analysis-show-potential-biden-election-landslide/?sh=41200bb7f0b6

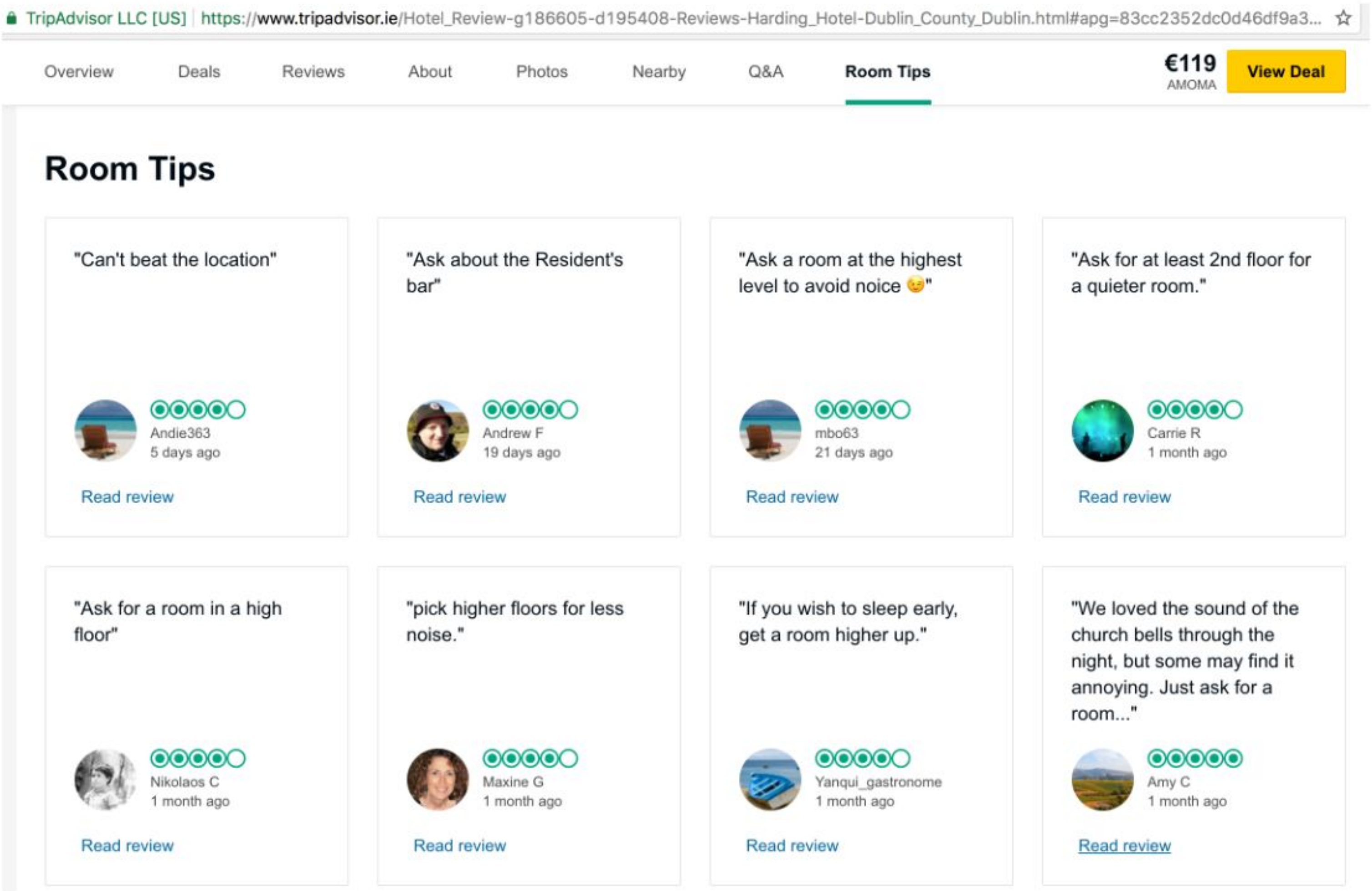
Sentiment Analysis in Finance



S&P 500 Index Description:

Index of the largest 500 US companies.

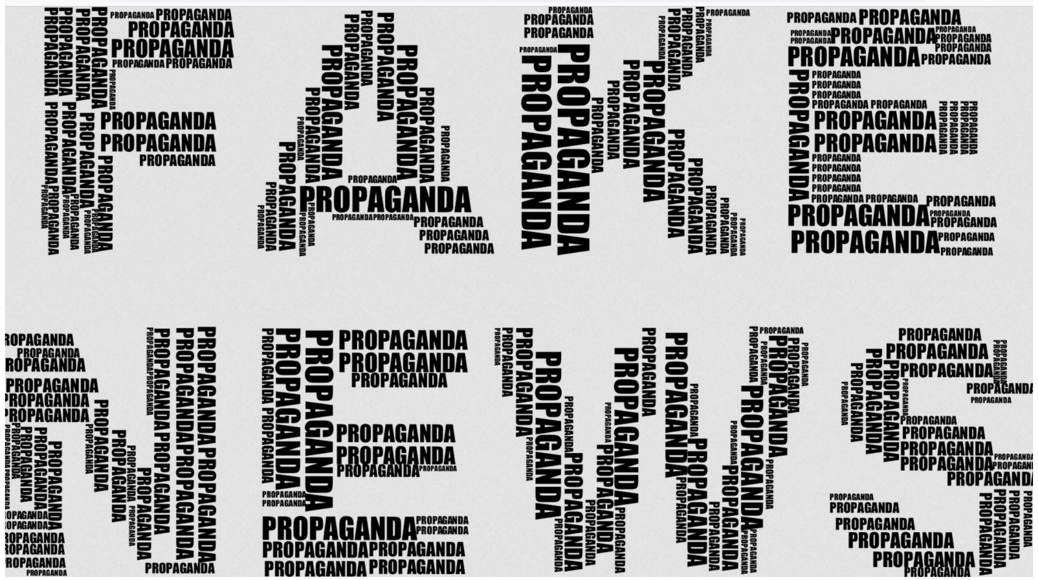
Suggestion Mining



Negi, S., 2019. *Suggestion mining from text.* PhD Thesis, NUI Galway

Other Social Media Analysis Applications

Offensive Content Identification



Misinformation Detection

https://www.mediafirst.co.uk/media/1697/offensive-resized.jpg
https://upload.wikimedia.org/wikipedia/commons/1/13/FAKE_NEWS.png

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

Buzz Groups

Can you think of any challenges in social media analysis?

Social Media Analysis Pipeline

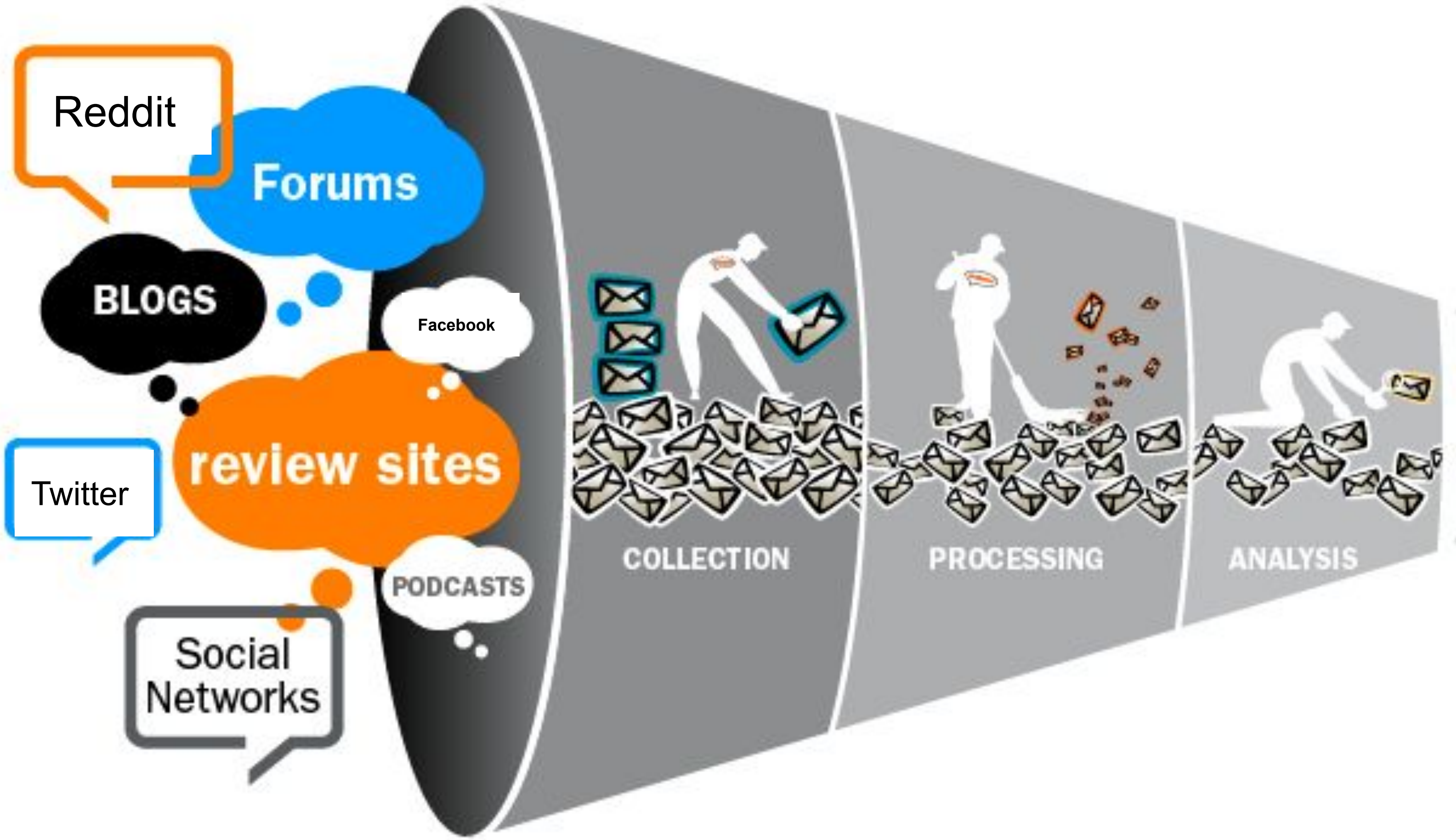


Image source https://www.promptcloud.com/wp-content/uploads/2015/05/Social-Media-Monitoring.png

Challenges - Data Collection

Copyright and redistribution policies

Data protection and privacy

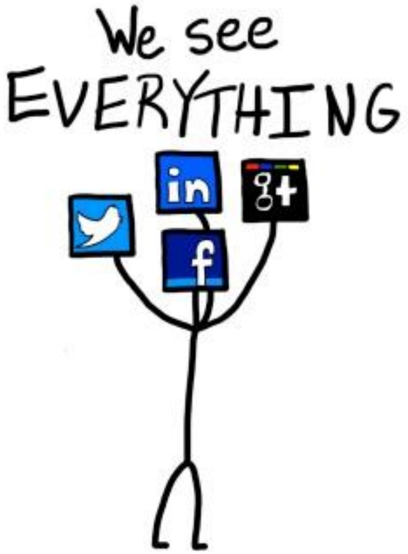
- GDPR (EU), other similar regulations

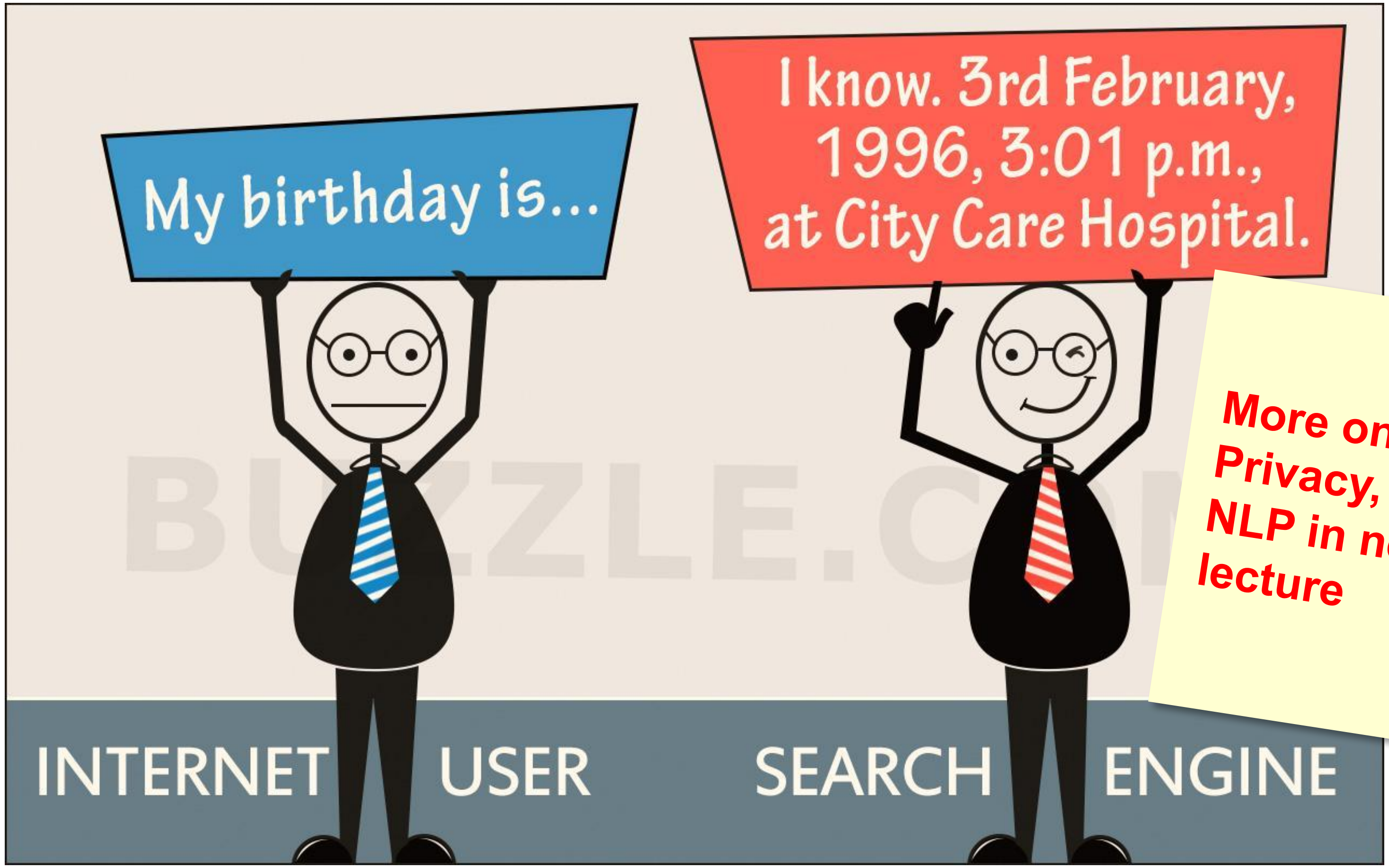
Ethical AI

- Gender and other bias
- Representative data

Image source https://www.kuppingercole.com/pics/2018_ai_banner_1000x563px.png

Data Collection – Ethical Considerations





More on Data Privacy, Ethics and NLP in next week lecture

Image source: buzzle.com

Data Collection – Sources

	Method	Effort
Already existing “benchmark” datasets	Download	LOW
Create new ones	APIs	MEDIUM
	Scrape/Crawl	HIGH

Important Note

Not every dataset is a benchmark

Important Note

Data scraping/crawling needs to be in line with legal guidelines

Data Collection – APIs

Data collection from social media, RSS feeds, blogs, wikis, news, …

Example: Twitter API

Twitter API

The Twitter API enables programmatic access to Twitter in unique and advanced ways. Use it to analyze, learn from, and interact with Tweets, Direct Messages, users, and other key Twitter resources.

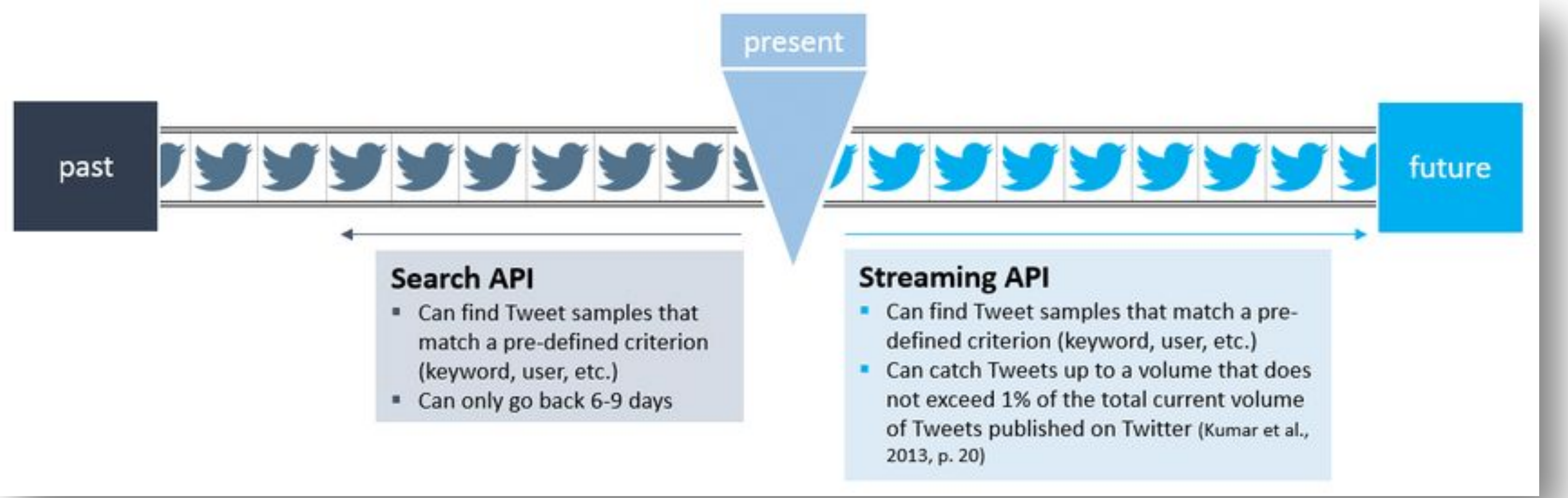


Image source: https://developer.twitter.com/en/docs/twitter-api

Challenges - Data Processing

Noise (urls, images, duplication)

Text normalisation

- Orthographic errors (**spelling**), non-standard language use (**abbreviations**)
- **Code-mixing** (different languages in one social media post)
- Design decisions on handling **punctuation, emojis, mentions, hashtags**

Some available tools: e.g. Tweet NLP, BERTweet

Tweet NLP
Carnegie Mellon

http://www.cs.cmu.edu/~ark/TweetNLP/

https://github.com/VinAIResearch/BERTweet

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

Sentiment Analysis

Sentiment Analysis Pipeline

Sentiment Analysis – Task Definition

Sentiment analysis input data

- Product reviews; Travel forums; Political discourse; Surveys; other!

Predict the sentiment expressed by input data

- **Binary Classification**: predict positive or negative sentiment class
- **Ordinal Classification**: predict a rate on a scale

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

Buzz Groups

Can you think of any challenges in Sentiment Analysis?

Challenges – Implicit Sentiment

Neutral words used but **+ve** sentiment implied

"Go read the book!"

Neutral words used but **-ve** sentiment implied

"If you are reading this because it is your darling fragrance, please wear it at home exclusively and tape the windows shut."

Challenges – Indirect Sentiment

Sentiment expressed may not be that of the author

*“**Although** this product is **disliked by many**, …”*

Challenges – Ambiguity

Same words used in different contexts with different sentiment

*"This car's steering is **unpredictable**!"* **-ve**

*"This film is **unpredictable**!"* **+ve**

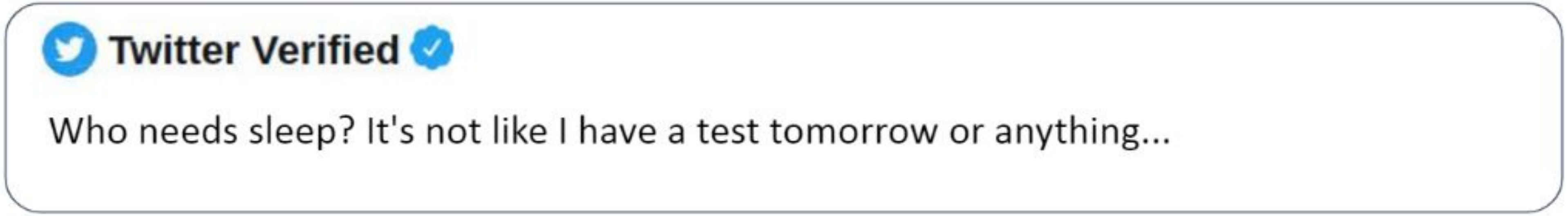
Examples of tweets in the following slides from: Pasricha, N. et al., 2021. *Sentiment Analysis: Message Polarity Classification.* Practical Project CRT-AI NLP Week, NUI Galway

Challenges – Irony & Sarcasm

Problem of determining if actual meaning of a word is intended

Sarcastic content could have **+ve** words but **-ve** sentiment implied

*“**Great job** Rogers! Raise rates but not service.”*



Challenges – Negation

+ve words used but **-ve** sentiment implied due to negation

*"I **don't like** this new Nokia model"*

Negation has diverse forms, so can be hard to detect:

*I **didn't** enjoy it.*

*I **never** enjoy it.*

***No one** enjoys it.*

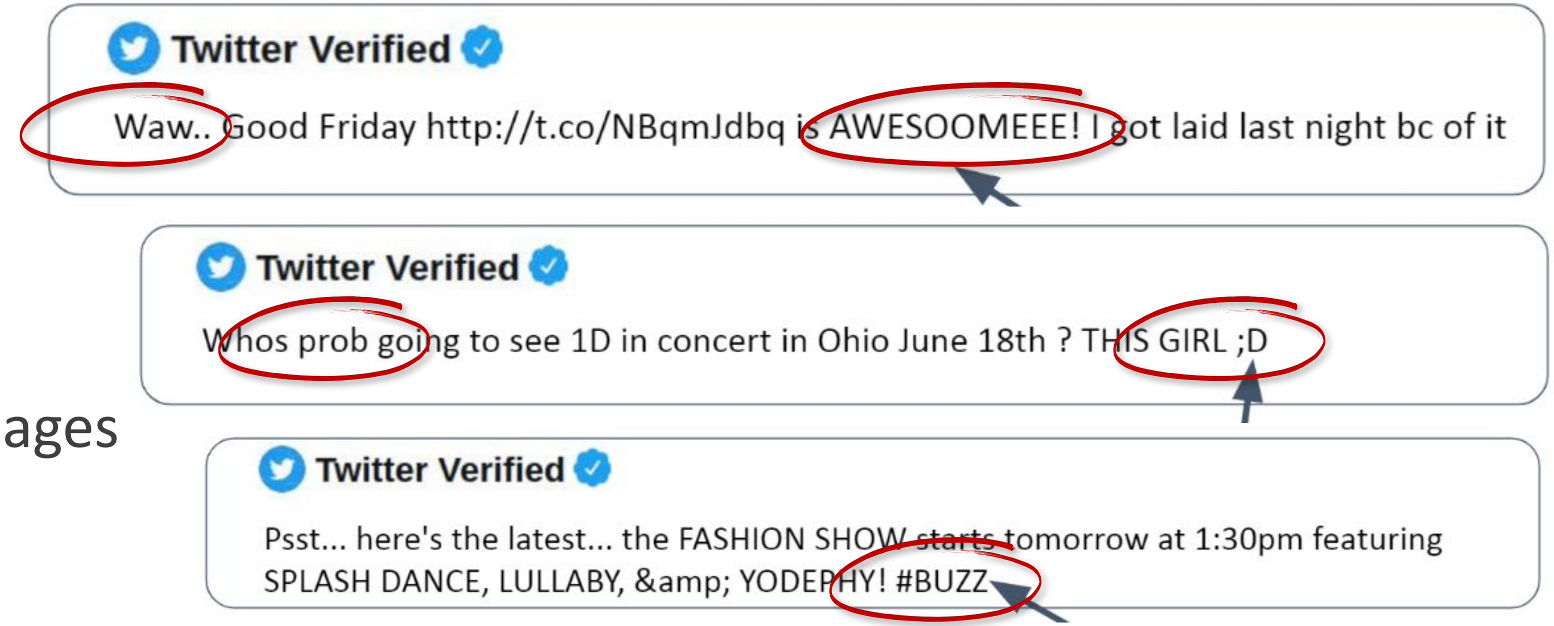
*I **have yet to** enjoy it.*

Challenges – Informal Language

Social media content uses non-standard, informal language

This includes the need to handle:

- Misspellings
- Emojis
- Slang
- Lack of context
- Hashtags, mentions, images



Overview of Sentiment Analysis Approaches

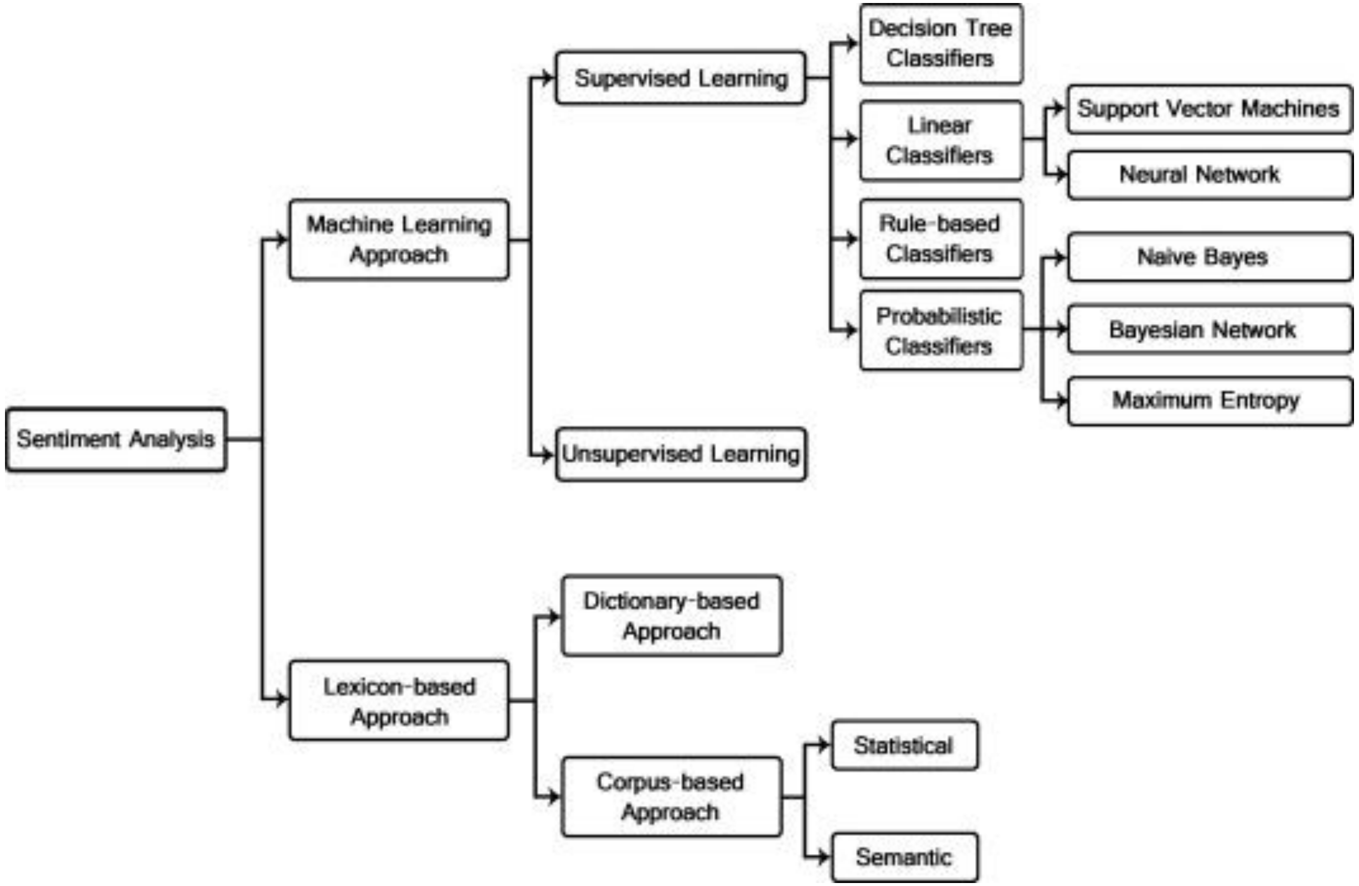


Image source: Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093-1113.

Sentiment Analysis Approaches

‘Unsupervised’ - Lexicon-based Approaches

- Lexicons – MPQA, LIWC, SentiWordNet
- VADER

Supervised - Probabilistic Approaches

- Machine Learning
  E.g. SVM, Naïve Bayes

Supervised - Deep Learning Approaches

- Neural Networks
  E.g. BERT/BERTweet

Require Annotated Data

Lexicon-based Approaches

IF there exists word *w* in sentence *S* such that *w* in *[good, great, nice, …]*,

THEN output **+ve**

Sentiment Lexicon

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

Sentiment Lexicon

Use of a Sentiment Lexicon

1. Calculate percentage of +ve/-ve words in text to be classified
2. Highest percentage determines sentiment

*"The camera's focus was **bad**, but has a **great** size and is **easy-to-use**."*

+ve: 2/13 = 0.153

-ve: 1/13 = 0.077

f (0.153, 0.077) = +ve

Sentiment Lexicon	
+ve	-ve
great	bad
superb	lonely
unlimited	disoriented
easy-to-use	hated
impeccably	horrible
fast-paced	awful
blessing	terrible
…	…

Sentiment Lexicon - Examples

- The General Inquirer (Stone et al. 1966) http://www.wjh.harvard.edu/~inquirer/
- MPQA subjectivity lexicon (Wilson et al. 2005) http://mpqa.cs.pitt.edu/lexicons/subj_lexicon
- AFINN lexicon (Nielsen 2011) http://corpustext.com/reference/sentiment_afinn.html
- NRC Word-Emotion Association Lexicon (EmoLex) (Mohammad & Turney 2013) https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm
- LIWC (Linguistic Inquiry and Word Count) Lexicon (Pennebaker 2015) https://www.liwc.app/
- SentiWordNet (Baccianella et al 2010) - https://github.com/aesuli/SentiWordNet

Sentiment Lexicon - MPQA

Examples from the MPQA Subjectivity Lexicon

Sr. No.	Strength	Length	Word	POS	Stemmed	Polarity
1	type= weaksubj	len=1	word1= abandoned	pos1= adj	stemmed1=n	priorpolarity = negative
2	type= weaksubj	len=1	word1= abandonment	pos1= noun	stemmed1=n	priorpolarity = negative
3	type= weaksubj	len=1	word1= abandon	pos1= verb	stemmed1=y	priorpolarity = negative
4	type= strongsubj	len=1	word1= abase	pos1= verb	stemmed1=y	priorpolarity = negative
5	type= strongsubj	len=1	word1= abasement	pos1= anypos	stemmed1=y	priorpolarity = negative
…						
8221	type= strongsubj	len=1	word1= zest	pos1= noun	stemmed1=n	priorpolarity = positive

Image from: R. S. Jagdale, V. S. Shirsat and S. N. Deshmukh, "Review on Sentiment Lexicons," *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*, 2018, pp. 1105-1110.

Sentiment Lexicon - SentiWordNet

Derived from WordNet, adding sentiment polarity on (words in) WordNet synsets

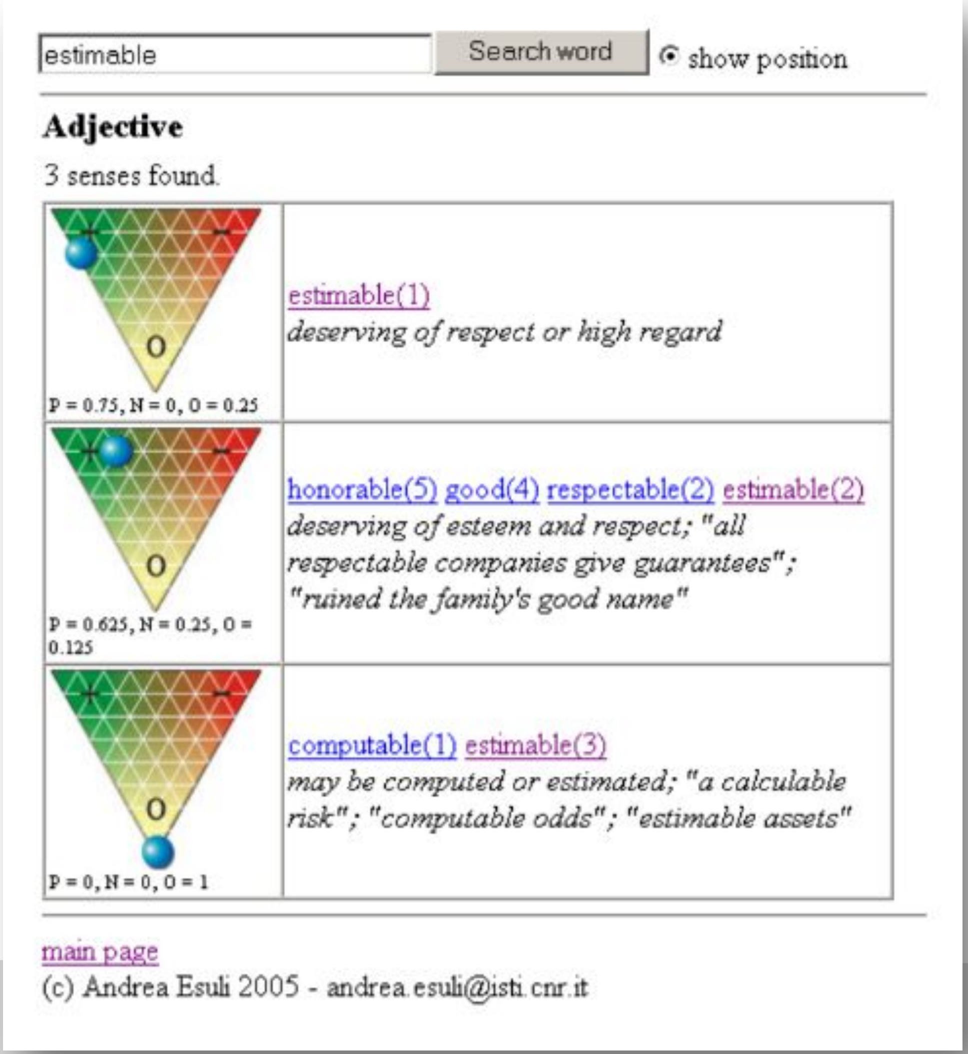
Table 1: The 10 top-ranked positive synsets and the 10 top-ranked negative synsets in SENTIWORDNET 3.0.

Rank	Positive	Negative
1	good#n#2 goodness#n#2	abject#a#2
2	better_off#a#1	deplorable#a#1 distressing#a#2 lamentable#a#1 pitiful#a#2 sad#a#3 sorry#a#2
3	divine#a#6 elysian#a#2 inspired#a#1	bad#a#10 unfit#a#3 unsound#a#5
4	good_enough#a#1	scrimy#a#1
5	solid#a#1	cheapjack#a#1 shoddy#a#1 tawdry#a#2
6	superb#a#2	unfortunate#a#3
7	good#a#3	inauspicious#a#1 unfortunate#a#2
8	goody-goody#a#1	unfortunate#a#1
9	amiable#a#1 good-humored#a#1 good-humoured#a#1	dispossessed#a#1 homeless#a#2 roofless#a#2
10	gainly#a#1	hapless#a#1 miserable#a#2 misfortunate#a#1 pathetic#a#1 piteous#a#1 pitiable#a#2 pitiful#a#3 poor#a#1 wretched#a#5

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of LREC.*

Sentiment Lexicon - SentiWordNet

Visualisation of sentiment properties of the adjective "*estimable*" across 3 senses (i.e. synsets)



Vader Algorithm / Tool

VADER (Valence Aware Dictionary for sEntiment Reasoning) uses a **curated lexicon derived from sentiment lexicons** that assigns a positivity/negativity score to 7k+ words/emoticons.

It also uses **hand-written pattern matching rules** (e.g., negation, intensifiers) to modify the contribution of the original word scores to the overall sentiment of text.

VADER is **integrated into NLTK**

```
NEGATE - {"aint", "arent", "cannot", "cant", "couldn
 "ain't", "aren't", "can't", "couldn't", "daren't",
 "dont", "hadnt", "hasnt", "havent", "isnt", "mightn
 "don't", "hadn't", "hasn't", "haven't", "isn't", "m
 "neednt", "needn't", "never", "none", "nope", "nor"
 "oughtnt", "shant", "shouldnt", "uhuh", "wasnt", "w
 "oughtn't", "shan't", "shouldn't", "uh-uh", "wasn't
 "without", "wont", "wouldnt", "won't", "wouldn't",

# booster/dampener 'intensifiers' or 'degree adverbs
# http://en.wiktionary.org/wiki/Category:English_deg

BOOSTER_DICT = \
{"absolutely": B_INCR, "amazingly": B_INCR, "awfully
 "decidedly": B_INCR, "deeply": B_INCR, "effing": B_
 "entirely": B_INCR, "especially": B_INCR, "exceptio
 "fabulously": B_INCR, "flipping": B_INCR, "flippin"
 "fricking": B_INCR, "frickin": B_INCR, "frigging":
 "greatly": B_INCR, "hella": B_INCR, "highly": B_INC
 "intensely": B_INCR, "majorly": B_INCR, "more": B_I
 "purely": B_INCR, "quite": B_INCR, "really": B_INCR
 "so": B_INCR, "substantially": B_INCR,
 "thoroughly": B_INCR, "totally": B_INCR, "tremendou
 "uber": B_INCR, "unbelievably": B_INCR, "unusually"
 "very": B_INCR,
 "almost": B_DECR, "barely": B_DECR, "hardly": B_DEC
 "kind of": B_DECR, "kinda": B_DECR, "kindof": B_DEC
 "less": B_DECR, "little": B_DECR, "marginally": B_D
 "scarcely": B_DECR, "slightly": B_DECR, "somewhat":
 "sort of": B_DECR, "sorta": B_DECR, "sortof": B_DEC

# check for special case idioms using a sentiment-la
SPECIAL_CASE_IDIOMS = {"the shit": 3, "the bomb": 3,
                       "cut the mustard": 2, "kiss o
```

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

Sentiment Analysis Approaches

‘Unsupervised’ - Lexicon-based Approaches

- Lexicons – MPQA, LIWC, SentiWordNet
- VADER

Supervised - Probabilistic Approaches

- Machine Learning
  E.g. SVM, Naïve Bayes

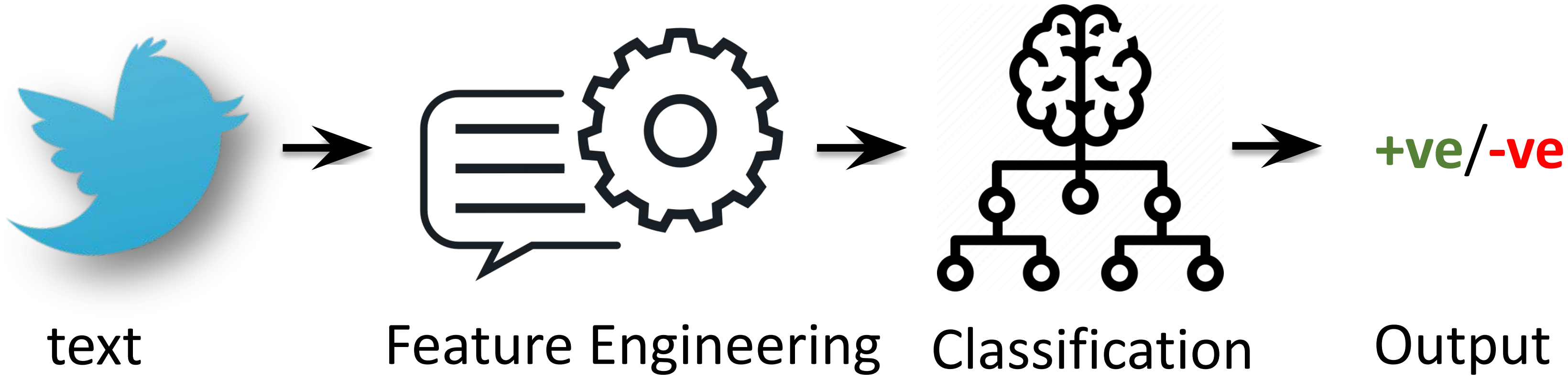
Supervised - Deep Learning Approaches

- Neural Networks
  E.g. BERT/BERTweet

Require Annotated Data

Probabilistic Approaches

Traditional ML methods can be used to classify a given text as **+ve**/**-ve**



Naïve Bayes
SVM
…

Feature Engineering

Sentiment features (sentiment lexicon)

Linguistic features

- N-grams
- part-of-speech (N, V, Adj, …)
- syntactic dependency (nsubj, dobj, amod, nmod, …)
- modal verbs (could be, should be, …)

Social Media features (hashtags, emoticons, …)

Other features (negation)

Feature selection and weighting: occurrence (binary), freq, PMI, TF-IDF

Sentiment Analysis – Approaches

‘Unsupervised’ - Lexicon-based Approaches

- Lexicons – MPQA, LIWC, SentiWordNet
- VADER

Supervised - Probabilistic Approaches

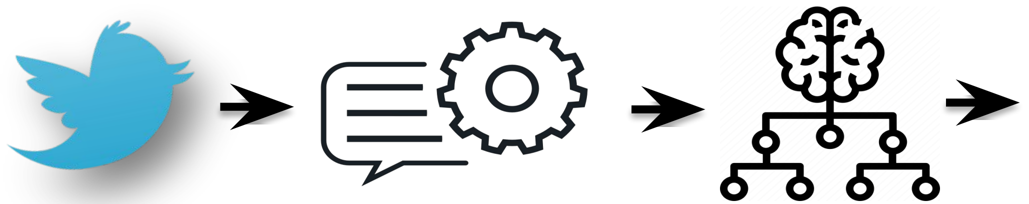
- Machine Learning
  E.g. SVM, Naïve Bayes

Supervised - Deep Learning Approaches

- Neural Networks
  E.g. BERT/BERTweet

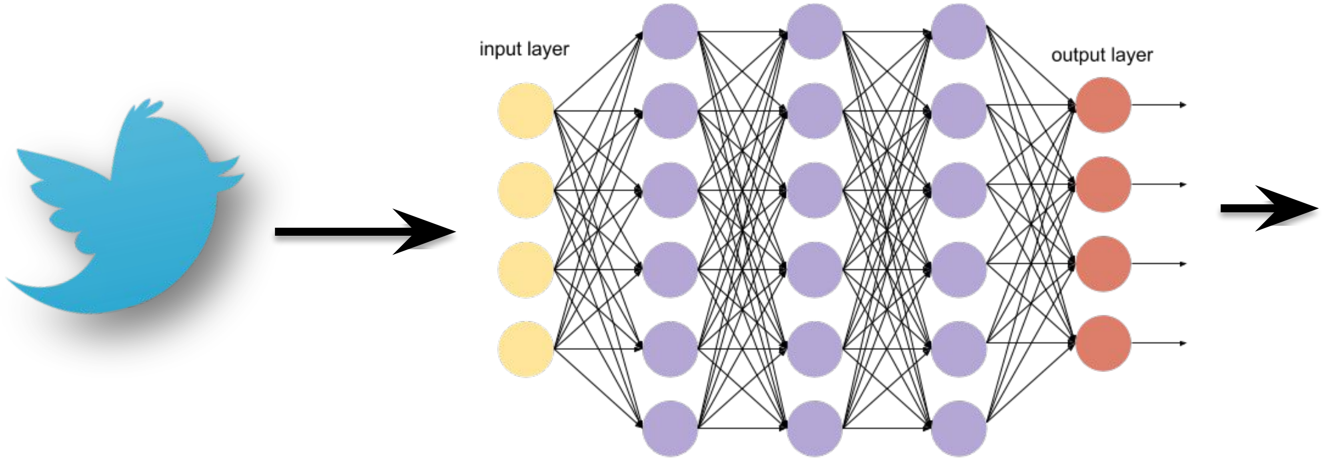
Require Annotated Data

From ML to Deep Learning Approaches



+ve/-ve

text Feature Engineering Classification Output



+ve/-ve

text

Feature **Learning** + Classification

Output

BERTweet

Pre-trained language model for English Tweets

Corpus used for pre-training:

- 850M English Tweets (16B word tokens ~ 80GB)
- 845M Tweets streamed from 01/2012 to 08/2019
- 5M Tweets related to the COVID-19 pandemic

Nguyen, D. Q., Vu, T., and Tuan Nguyen, A. (2020). BERTweet: A pre-trained language model for English tweets. In Proceedings of Empirical Methods in Natural Language Processing: System Demonstrations, pp 9–14.

https://github.com/VinAIResearch/BERTweet

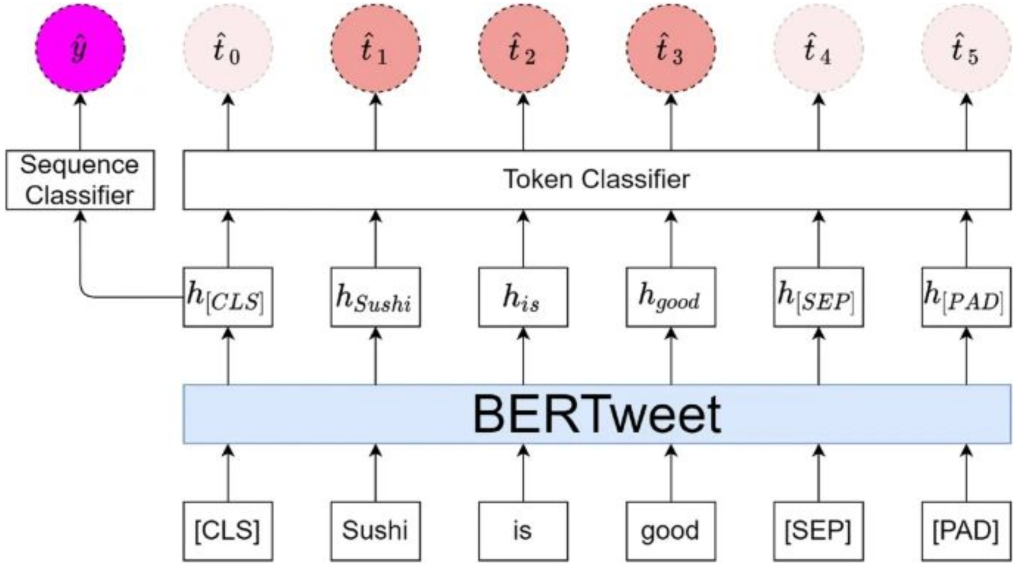
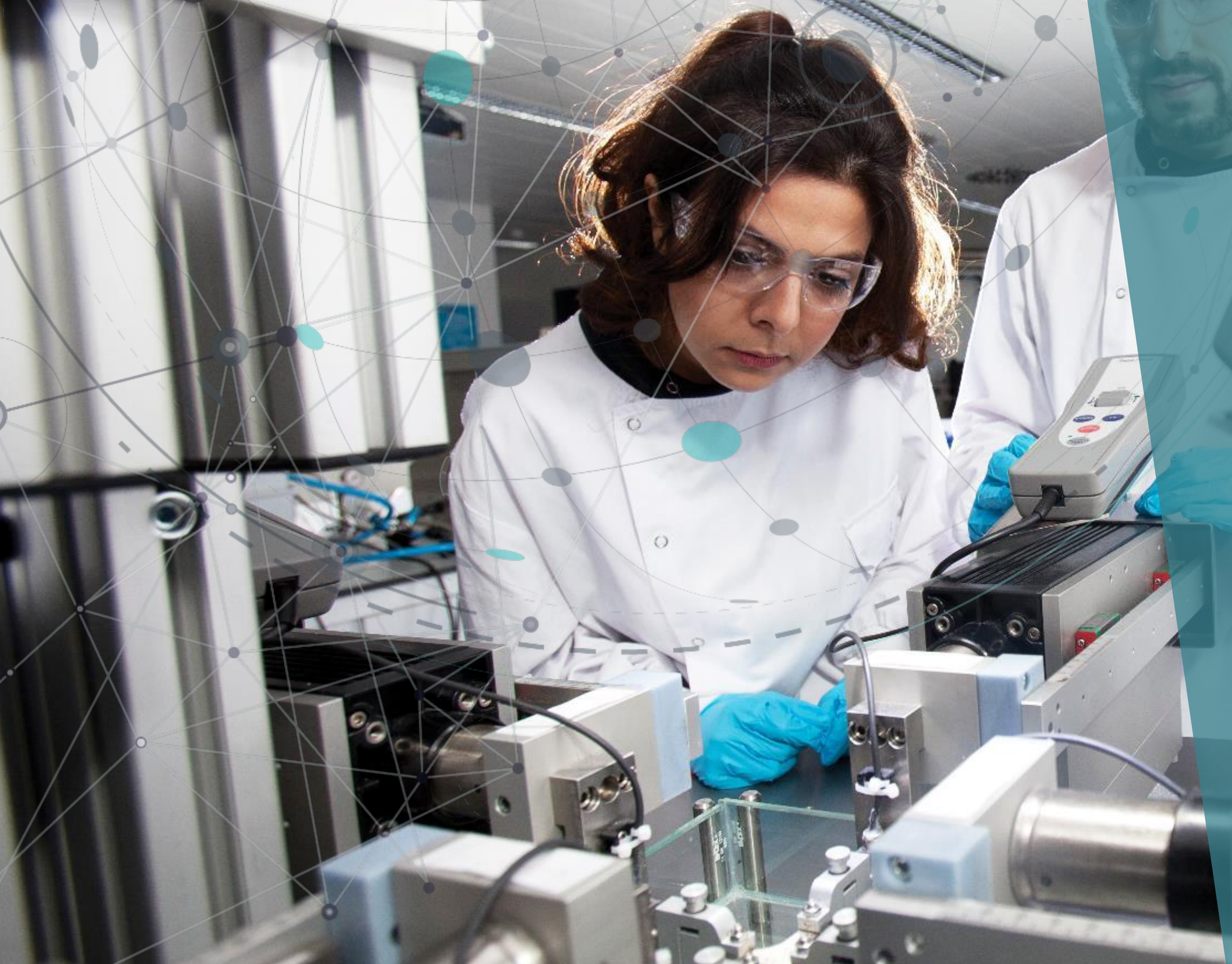


Image source: Tao, D., Zhang, D., Hu, R., Rundensteiner, E., & Feng, H. (2021). Crowdsourcing and machine learning approaches for extracting entities indicating potential foodborne outbreaks from social media. *Scientific reports*, *11*(1), 1-12.

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY



Buzz Groups

What is missing sofar in the discussion of sentiment analysis?

Aspect-based Sentiment Analysis

Polarity: Rate the sentiment of the text as +ve or –ve

*“The camera’s focus was **bad**, but has a **great** size and is **easy-to-use**.”*

Aspect: Detect the sentiment **aspect** (‘target/topic of the sentiment’)

*“The camera’s focus was **bad**, but has a **great** size and is **easy-to-use**.”*

Multiple Aspects

*“The **staff** was very friendly, the **room** was comfortable and **breakfast** was tasty, but the **bathroom** was very small.”*

Aspect	**Sentiment**
Staff	**+ve**
Breakfast	**+ve**
Room	**+ve**
Bathroom	**-ve**

Identifying the Aspect(s)

Supervised Method

- Manually define a (small) set of aspect categories, e.g. *'staff, breakfast, room, …'* for the hotel domain
- Construct a labeled dataset with aspects and corresponding sentiments

  *"The [**staff** was very friendly], the [**room** was comfortable] and [**breakfast** was tasty], but the [**bathroom** was very small]."*

- Build a multi-label classifier which assigns aspects + sentiments to a sentence

  staff +ve, room +ve, breakfast +ve, bathroom -ve

Identifying the Aspect(s)

Semi-supervised Method

- Manually define a (small) set of aspect categories, e.g. *'staff, breakfast, room, …'* for the hotel domain
- Check for these aspect categories in a window around sentiment words from a sentiment lexicon
- Compute sentiment for aspect categories in window

*"The **staff** was very friendly, the room was comfortable and breakfast was tasty, but the bathroom was very small."*

Identifying the Aspect(s)

Semi-supervised Method

- Manually define a (small) set of aspect categories, e.g. *'staff, breakfast, room, …'* for the hotel domain
- Check for these aspect categories in a window around sentiment words from a sentiment lexicon
- Compute sentiment for aspect categories in window

*"The staff was very friendly, the **room** was comfortable and breakfast was tasty, but the bathroom was very small."*

Sentiment Analysis - Evaluation

Evaluate on labeled data (‘ground truth’, ‘gold standard’)

- **True Positive (TP)**: Item is of class A (pos/neg), model assigned class A.
- **True Negative (TN)**: Item is not of class A, model assigned a class different from A.
- **False Positive (FP)**: Item is of class different of A, model assigned class A.
- **False Negative (FN)**: Item is of class A, model assigned a class different from A.

Precision = TP/(TP+FP)

Recall = TP/(TP+FN)

Accuracy = (TP+TN)/(TP+TN+FP+FN)

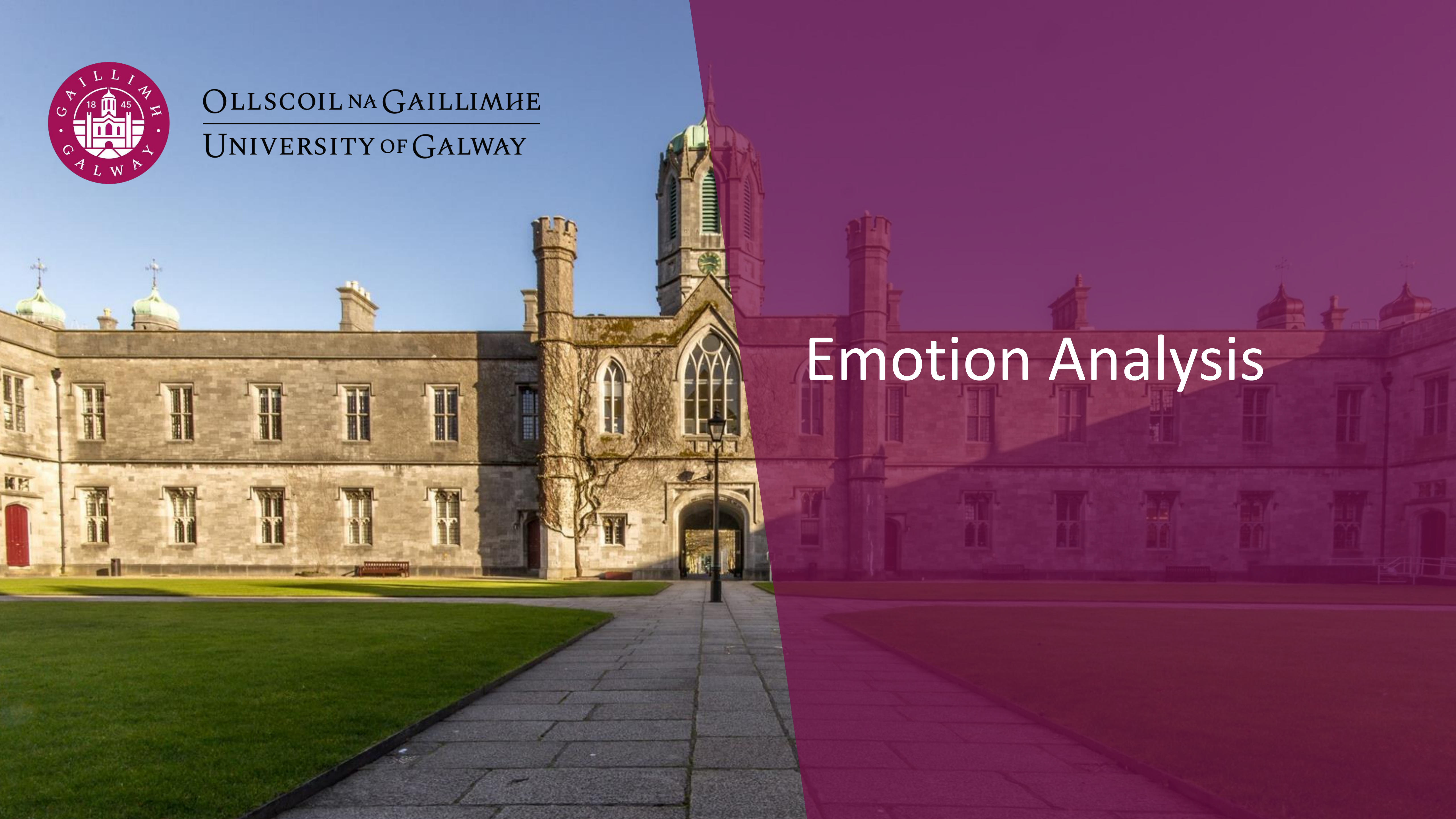
https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/

Sentiment Analysis - Summary

Consider the following in building a sentiment classifier

- What is the input for each prediction? (sentence; paragraph; document)
- What are the possible outputs? (+ve/−ve; scale; aspect)
- What will be the used approach? (rule-based; traditional ML; DL)
- How will you evaluate the system? (ground truth; precision, recall, accuracy)

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY



Emotion Analysis

Scherer Typology of Affective States

Emotion

- angry, sad, joyful, fearful, ashamed, proud, elated

Mood

- cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stances

- friendly, flirtatious, distant, cold, warm, supportive, contemptuous

Attitudes

- liking, loving, hating, valuing, desiring

Personality traits

- nervous, anxious, reckless, morose, hostile, jealous

Scherer, K. R., Dan, E., & Flykt, A. (2006). What determines a feeling's position in three-dimensional affect space?. *Cognition Emotion*, *20*(1), 92-113.

Scherer Typology - Sentiment Analysis

Emotion

- angry, sad, joyful, fearful, ashamed, proud, elated

Mood

- cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stances

- friendly, flirtatious, distant, cold, warm, supportive, contemptuous

Attitudes

- **liking, loving, hating, valuing, desiring**

Personality traits

- nervous, anxious, reckless, morose, hostile, jealous

Scherer, K. R., Dan, E., & Flykt, A. (2006). What determines a feeling's position in three-dimensional affect space?. *Cognition Emotion*, *20*(1), 92-113.

Scherer Typology - Emotion Analysis

Emotion

- **angry, sad, joyful, fearful, ashamed, proud, elated**

Mood

- cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stances

- friendly, flirtatious, distant, cold, warm, supportive, contemptuous

Attitudes

- liking, loving, hating, valuing, desiring

Personality traits

- nervous, anxious, reckless, morose, hostile, jealous

Scherer, K. R., Dan, E., & Flykt, A. (2006). What determines a feeling's position in three-dimensional affect space?. *Cognition Emotion*, *20*(1), 92-113.

Sentiment Analysis vs. Emotion Analysis

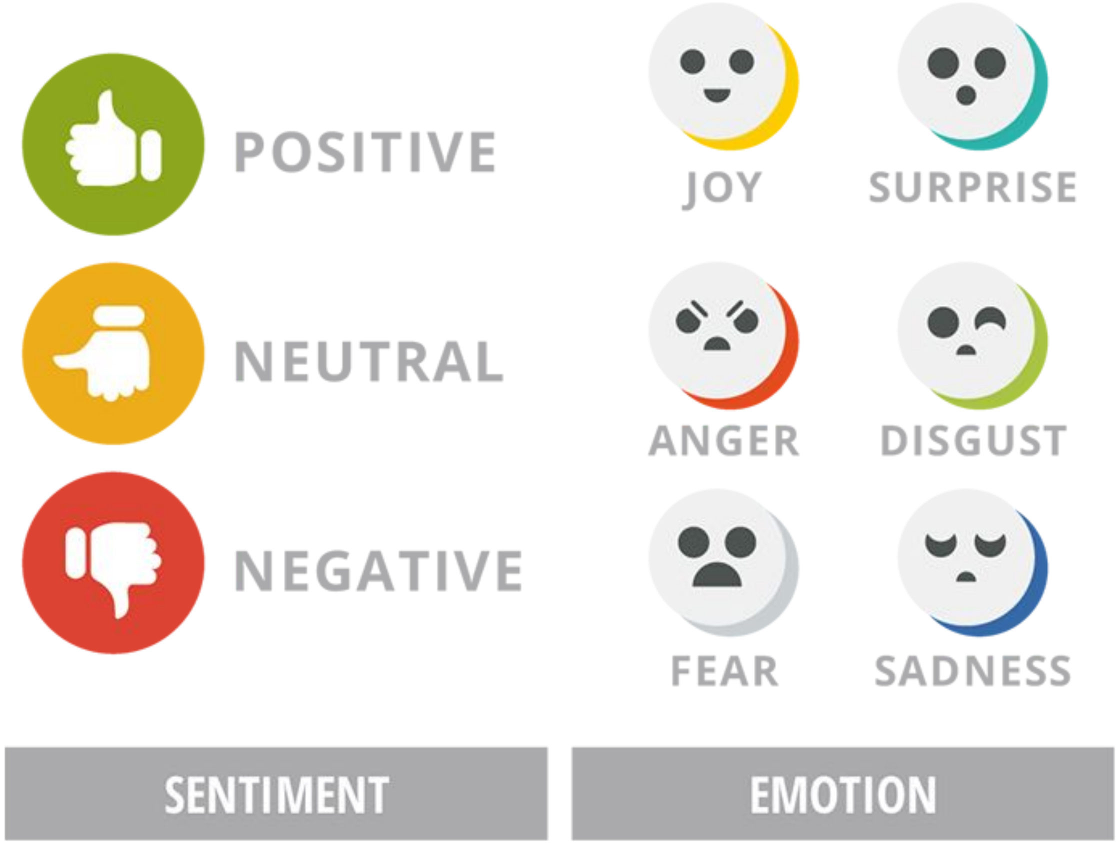
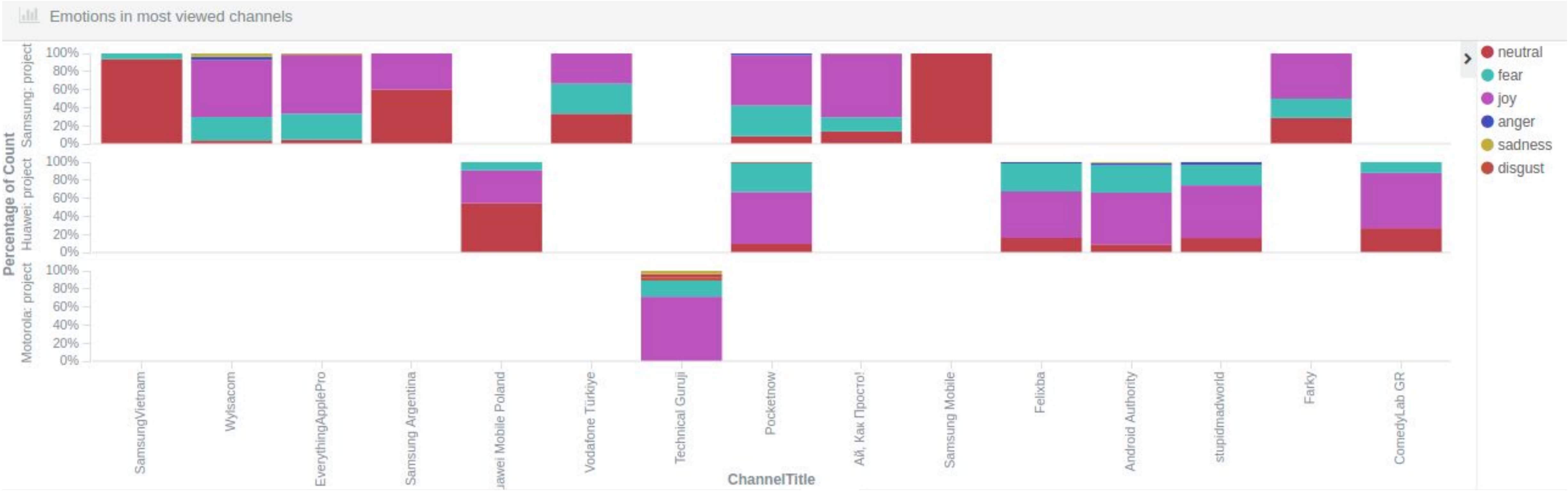


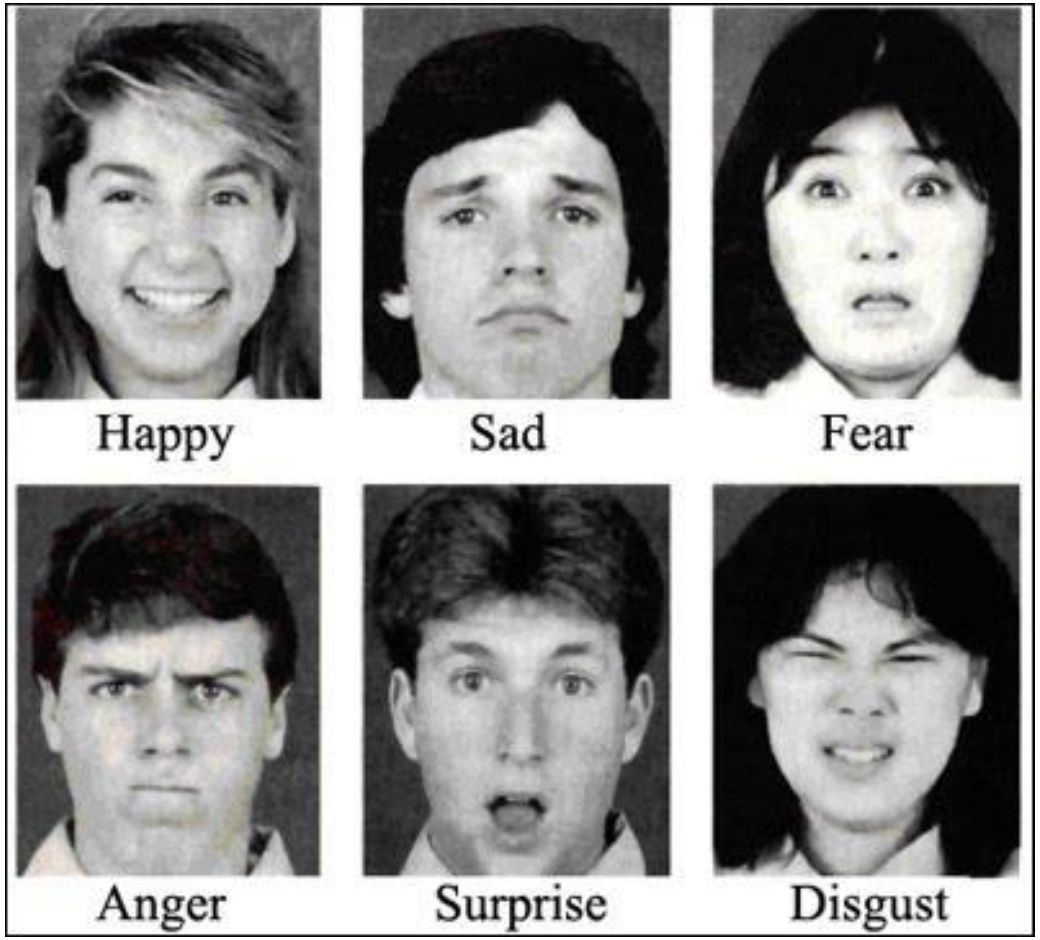
Image source
https://www.researchgate.net/publication/331523338_Improve_the_User_Experience_of_Chatbots_with_Tone_Analysis

Emotion Analysis - Use Case



Emotion Models - Ekman

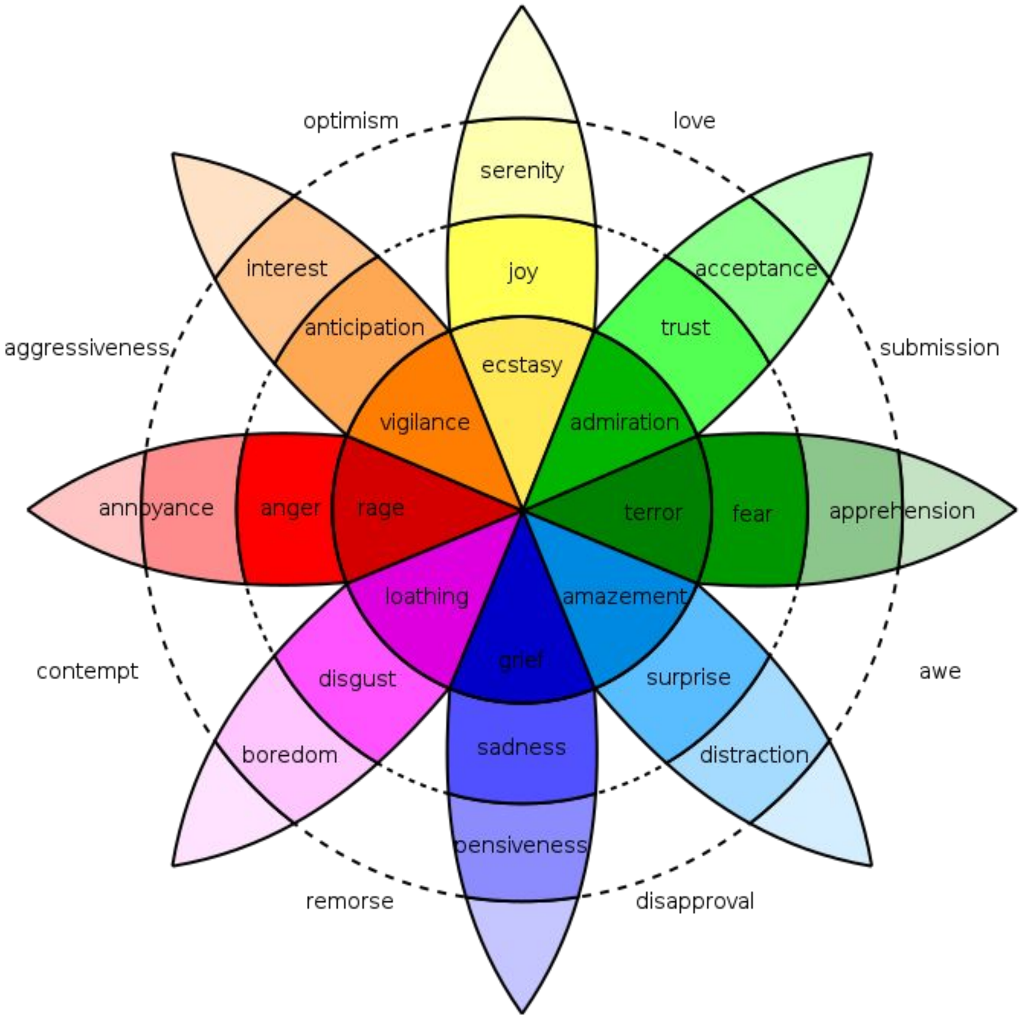
Most widely used model is **Ekman's** with **6 basic facial emotions**



Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*(4), 384–392.

Emotion Models - Plutchik

‘Plutchik’s wheel’ has 2 more categories (Trust, Anticipation) and 4 levels of intensity to give **32 emotions**

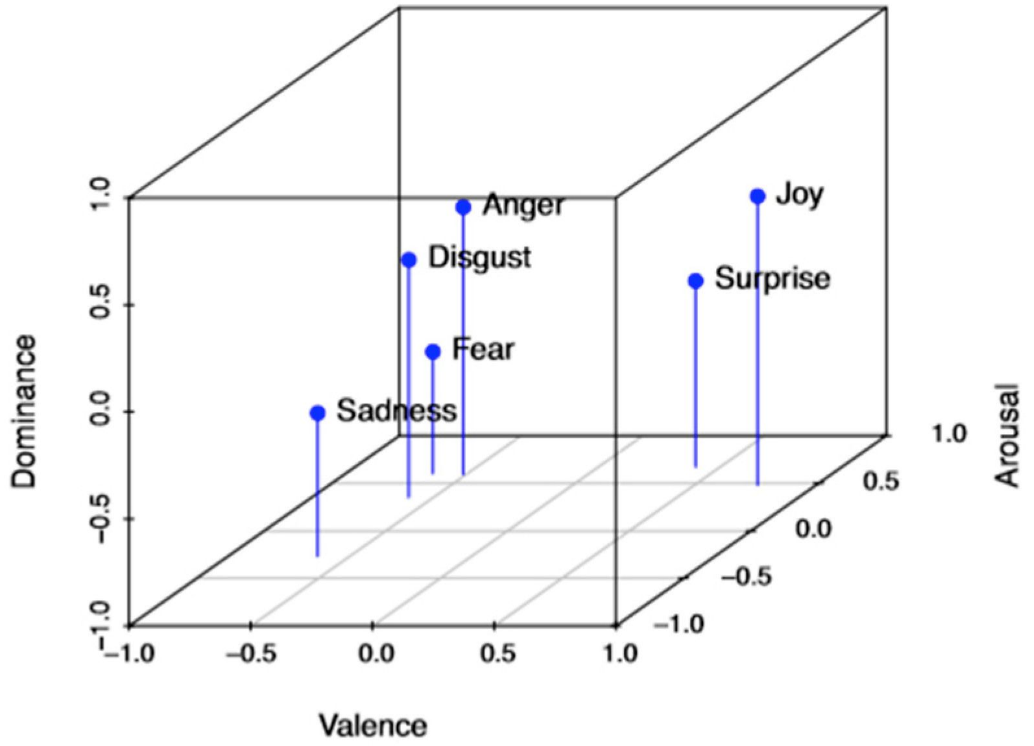


Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *Theories of emotion* (pp. 3-33). Academic press.

Emotion Models - Lövheim

Lövheim's model uses three continuous values related to neurotransmitters that have been interpreted as Valence, Arousal, Dominance

Emotions can be expressed by a combination of continuous values for V/A/D



Lövheim, H. (2012). A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical hypotheses*, *78*(2), 341-348.

Image from: Bălan, O., Moise, G., Petrescu, L., Moldoveanu, A., Leordeanu, M., & Moldoveanu, F. (2019). Emotion classification based on biophysical signals and machine learning techniques. *Symmetry*, *12*(1), 21.

Emotion Analysis Dataset - Example

Dataset Preview API

Subset: subtask5.english

Split: train

ID (string)	Tweet (string)	anger (bool)	anticipation (bool)	disgust (bool)	fear (bool)	joy (bool)	love (bool)	optimism (bool)	pessimism (bool)	sadness (bool)	surprise (bool)	trust (bool)
"2017-En-21441"	"“Worry is a down payment on a problem you may never have'.…	false	true	false	false	false	false	true	false	false	false	true
"2017-En-31535"	"Whatever you decide to do make sure it makes you #happy."	false	false	false	false	true	true	true	false	false	false	false
"2017-En-21068"	"@Max_Kellerman it also helps that the majority of NFL coachin…	true	false	true	false	true	false	true	false	false	false	false
"2017-En-31436"	"Accept the challenges so that you can literally even feel the…	false	false	false	false	true	false	true	false	false	false	false
"2017-En-22195"	"My roommate: it's okay that we can't spell because we have…	true	false	true	false	false	false	false	false	false	false	false
"2017-En-22190"	"No but that's so cute. Atsu was probably shy about photos before…	false	false	false	false	true	false	false	false	false	false	false
"2017-En-20221"	"Do you think humans have the sense for recognizing impending…	false	true	false	false	false	false	false	true	false	false	false

Mohammad, S., Bravo-Marquez, F., Salameh, M., & Kiritchenko, S. (2018). Semeval-2018 task 1: Affect in tweets. In *Proceedings of the 12th international workshop on semantic evaluation* (pp. 1-17).
https://huggingface.co/datasets/sem_eval_2018_task_1/viewer/subtask5.english/train

Ollscoil na Gaillimhe
University of Galway

Offensive Content Identification

Offensive Content Identification

Objective: automatically identify *offensive content*, also referred to as *hate speech* or (in NLP) *toxic spans*

Definition of hate speech by 'Global Handbook on Hate Speech Laws':

"...hate speech is understood as any form of negative speech (ranging from offensive to direct incitement to violence) targeted at protected characteristics such as sexual orientation, race or disability…"

https://futurefreespeech.com/global-handbook-on-hate-speech-laws/

Toxic Spans Detection - SemEval-2021

OFFSETS	POST
{}	I will stand by my statement that liberalism is what smart looks like to stupid people. Th yesterday states the real problem clearly enough. "It doesn't matter if it was protecte Amendment." I have legal news flash for the University of Oregon. Because the Universi receives government tax dollars it is indeed bound to abide by the US constitution wi PROTECTION INCLUDED. Which should tell everyone right off the signatories are and perhaps even anti-America in their twisted PC nonsense.
{}	Fits the despicable piece of trash like a glove.
{}	No he shot someone so he shouldn't be free. He should be locked up like He is for You holla like its no big deal but what if it was your little brother who was shot? Th wouldnrlt be screaming free ace g or whatever gay nickname he has. That is the lam ever. Do you think that name scares people? Ace clown is what i will now call him.
{55, 56, 57, 58}	Time to give the opposition weapons and overthrow this jerk.
{31, . . . , 39}	you are a paid liberal hack of stupidity shore living the liberal lie every day
{9, . . . , 13}	Youre an idiot.
{7, 8, . . . , 11, 14, 15, . . . , 18}	what a dummy, dummy has no regard for the law. you cannot brutalize a suspect. he has complete lack of respect for any law and is acting like a dictator. he is trying to emulate putin.
{12, . . . , 17, 94, . . . , 102}	People make stupid decisions and then expect the gov't to bail them out. There is no cure for stupidity.
{14, . . . , 20, 29, . . . , 35}	Nah, the only asshole is the asshole firing a rifle within city limits.

DISCLAIMER: These examples contain offensive wording, which are only used for illustration here and in the original paper.

Pavlopoulos, J., Sorensen, J., Laugier, L., & Androutsopoulos, I. (2021). Semeval-2021 task 5: Toxic spans detection. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)* (pp. 59-69).

Toxic Span SemEval-2021 - Best System

HITSZ-HLT at SemEval-2021 Task 5: Ensemble Sequence Labeling and Span Boundary Detection for Toxic Span Detection

Qinglin Zhu[1,†], Zijie Lin[1,†], Yice Zhang[1,†], Jingyi Sun[1], Xiang Li[1]
Qihui Lin[1], Yixue Dang[2], Ruifeng Xu[1,‡]
[1]Joint Lab of HITSZ-CMS, Harbin Institute of Technology(Shenzhen), China
[2]China Merchants Securities Co., Ltd

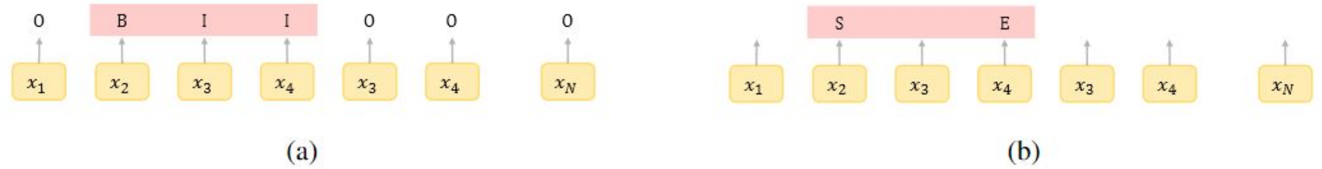


Figure 1: Comparison of SL and SBD, (a) denotes SL, (b) denotes SBD.

Zhu, Q., Lin, Z., Zhang, Y., Sun, J., Li, X., Lin, Q., ... & Xu, R. (2021, August). HITSZ-HLT at SemEval-2021 task 5: Ensemble sequence labeling and span boundary detection for toxic span detection. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)* (pp. 521-526).

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

Misinformation Detection

SemEval-2021 Task 6 - ‘Persuasion in Memes’

Subtask 1: Given the textual content of a meme, **identify which techniques (out of 20 possible ones) are used** in it. This is a **multilabel classification** problem.

Subtask 2: Given the textual content of a meme, **identify which techniques (out of 20 possible ones) are used** in it together with the span(s) of text covered by each technique. This is a **multilabel sequence tagging** task.

Subtask 3: Given a meme, **identify which techniques (out of 22 possible ones) are used** in the meme, **considering both the text and the image**. This is a **multilabel classification** problem.

Dimitrov, Dimitar, et al. "SemEval-2021 task 6: detection of persuasion techniques in texts and images." *arXiv preprint arXiv:2105.09284* (2021).

SemEval-2021 Task 6 - Data

DISCLAIMER: The data set example contains offensive wording, which is only used for illustration and instruction purposes here


```
{
        "id": "46_batch_2",
        "labels": [
            "Loaded Language",
            "Appeal to authority",
            "Misrepresentation of Someone's Position (Straw Man)"
        ],
        "text": "\"The phrase Black Lives Matter suggests racial superiority.\nIt excludes the importance of anybody else's life.\"\n\"THAT IS RACISM DEFINED.\"\nSHERIFF DAVID CLARKE\n"
    },
```

SemEval-2021 Task 6 - Data

DISCLAIMER: The data set example contains offensive wording, which is only used for illustration and instruction purposes here

{
    "id": "172_batch_2",
    "labels": [
        "Appeal to fear/prejudice",
        "Misrepresentation of Someone's Position (Straw Man)",
        "Loaded Language",
        "Name calling/Labeling"
    ],
    "text": "BILL GATES HAS A SECRET EUGENICS PLAN TO DEPOPULATE THE WORLD THROUGH VACCINES!\n\nBECAUSE ALL EVIL VILLAINS ANNOUNCE THEIR MURDEROUS PLANS IN TED TALK PRESENTATIONS."
},

SemEval Tasks (Ideas for MSc Projects)

SemEval-2016 Task 5: Aspect-Based Sentiment Analysis: https://alt.qcri.org/semeval2016/task5/

SemEval-2018 Task 1: Affect in Tweets (emotion analysis): https://competitions.codalab.org/competitions/17751

SemEval-2019 Task 5: Multilingual Detection of Hate Speech: https://competitions.codalab.org/competitions/19935

SemEval-2021 task 6 on Detection of Persuasion Techniques in Texts and Images: https://github.com/di-dimitrov/SEMEVAL-2021-task6-corpus

Learning Outcomes

After completing this lecture, you should be able to:

- Understand and apply baseline approaches for sentiment analysis
- Understand data labeling and task definition for emotion analysis, offensive content identification, and misinformation detection

This lecture is based on parts of:

Chapter 20 in Jurafsky and Martin, SPEECH and LANGUAGE PROCESSING: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 3rd edition: https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Lab of this week

Practical applications of sentiment analysis