Natural Affect Detection (Nade): Using Emojis to Infer Emotions from Text

European Quant Marketing Workshop

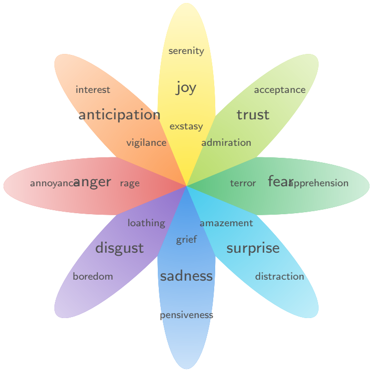
Christian Hotz-Behofsits Nils Wlömert Nadia Abou Nabout

Vienna University of Economics and Business

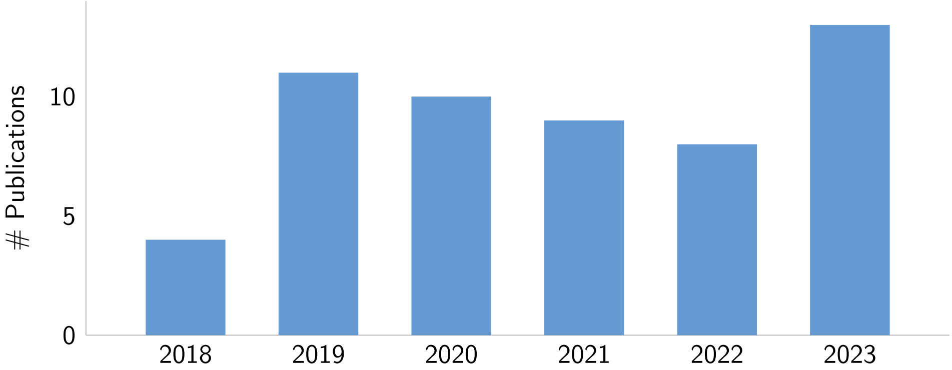
May 5$^{\text{th}}$, 2025

Extracting emotions from text

- Sentiment analysis has become very popular in research & practice
- Relationship between emotion and social transmission more **complex than valence alone** (Kranzbühler et al., 2020):
  - *Angry* reviews are considered less helpful than sad ones (Yin, Bond, and Zhang, 2014)
  - *Fearful* and *angry* newspaper articles go viral more than *sad* articles (Berger and Milkman, 2012)
  - *Anxious* and *angry* language holds attention in news consumption, while *sad* language discourages it (Berger, Moe, and Schweidel, 2023)
  - Social media users show differential emotional reactions to fake news (Vosoughi, Roy, and Aral, 2018)



Emotion Extraction in Marketing Research[1]



Only **31%** go beyond basic sentiment analysis.

100% of the articles that go beyond basic sentiment use dictionary-based methods.

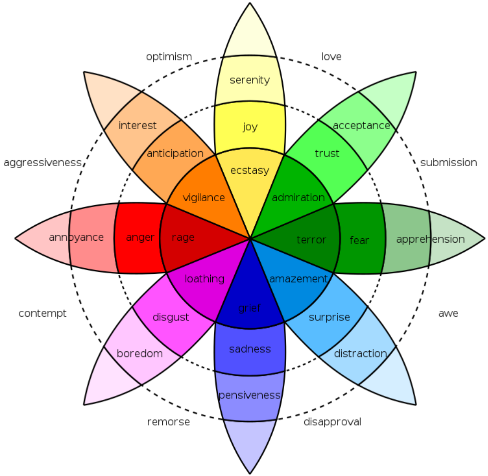
[1] Top-5 Journals, Six years from 2018 to 2023

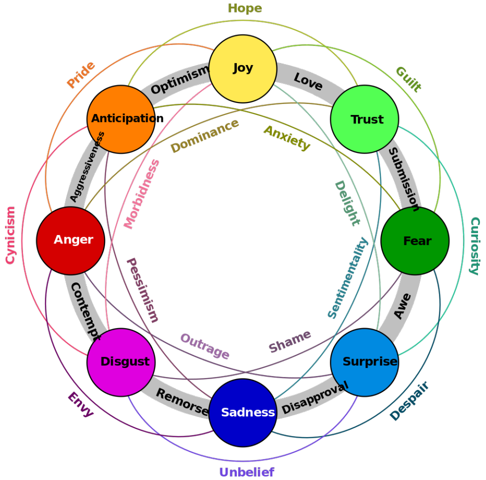
Emotion theories

- **Discrete Emotions** (Ekman, 1992)
  - Basic, universally recognized emotions (e.g., anger, disgust, fear, joy, sadness, and surprise).
  - Core assumption: Distinct emotions with universal expression and recognition.
- **Dimensional Systems** (Russell, 2003)
  - Emotions represented in a continuous space by dimensions (valence, arousal, dominance).
  - Core assumption: Emotions can be understood through general dimensions, not specific categories.
- **Hybrid Models** (Plutchik, 1980)
  - Integrates discrete and dimensional theories for nuanced understanding.
  - Core assumption: Emotions are complex, combining discrete categories with dimensional intensity and interactions.

Advantages of Plutchik's framework: (1) Conceptualizes emotional intensities, (2) comprises a broad range of 32 emotions, (3) easy to interpret.

Plutchik's Wheel of Emotions





Source: https://en.wikipedia.org/wiki/Emotion_classification

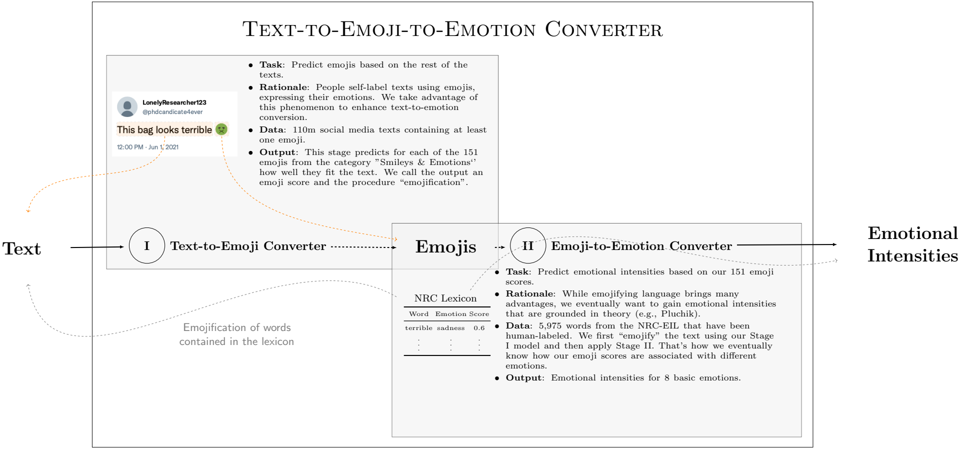
Measuring emotions

- **Human raters**
  - *Advantages:* Intuitive understanding of emotions; no need for technical expertise.
  - *Disadvantages:* Time-consuming and costly; not scalable for large datasets.
- **Lexicon/dictionary-based**
  - *Advantages:* Accessible and easy to use; low computational requirements.
  - *Disadvantages:* Limited vocabulary; costly to develop; struggles with slang, typos and new expressions.
- **Machine Learning (ML)-based**
  - *Advantages:* Scalable; can handle complex patterns in data.
  - *Disadvantages:* Requires large annotated datasets; may overlook nuances in emotions.
- **Large Language Models (LLMs)**
  - *Advantages:* Leverages vast unannotated data; reduces need for pre-labeling; advanced understanding of context.
  - *Disadvantages:* Insufficient validation; needs fine-tuning; privacy concerns.[2][3]

[2] https://www.wired.com/story/the-generative-ai-search-race-has-a-dirty-secret/

[3] https://www.bloomberg.com/news/articles/2023-03-09/how-much-energy-do-ai-and-chatgpt-use-no-one-knows-for-sure

Nade architecture



Stage I: Classifier

Aim Encode text to emojis

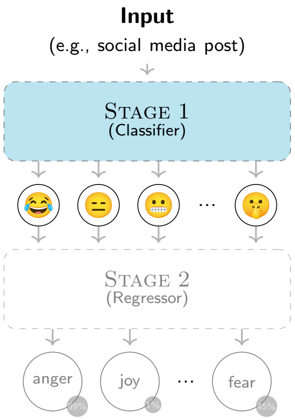
Approach Train a classifier to predict observed emojis

Data 110 million "clean"[4] English tweets

- 151 emojis from the category *smileys & emotion* (e.g., 💙, 😂)
- User generated content that comprises slang, typos, etc.

Model Efficient text classifier[5]

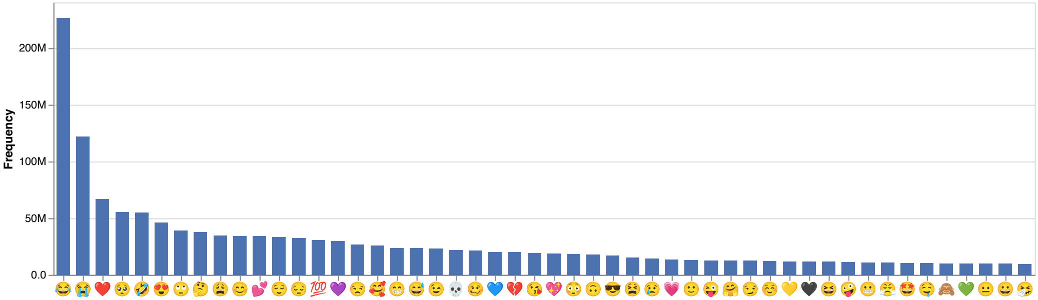
- FastText (Joulin et al., 2016)
- Handles unknown words and typos by exploiting subword information (Bojanowski et al., 2017)



[4] Unique, w/o hyperlink or media, single English sentence, and at least one relevant emoji.

[5] The trained model builds on a vocabulary of 9,796 (sub-)words.

Distribution of emojis

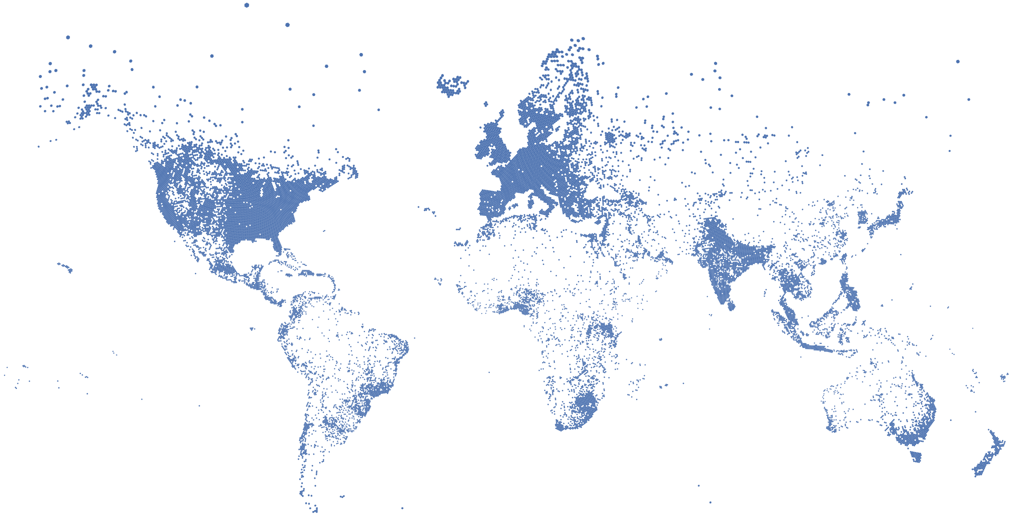


Emoji prediction results

Sentence	#1	#2	#3	#4	#5
love girls who have a good sense of music and amazon prime	❣️	❤️	💟	💙	💯
Amazon music is my new fav thing omg how did I not know about this before	🤯	😮	🙀	🙀	😯
i love amazon shopping	🤑	☺️	💋	😝	🤪
It used to be fun shopping on amazon	😅	☹️	🙃	☺️	🤣
Waiting impatiently for my box to arrive from Amazon Prime	🙃	😅	☺️	😤	🥲
Im gonna miss Amazon so much	☹️	😿	😞	😢	😣
amazon has been a real piece of shit lately	💩	🤬	👿	😡	🤥
Amazon sucks where are my packages ??????	🤬	😡	😒	😠	🤨

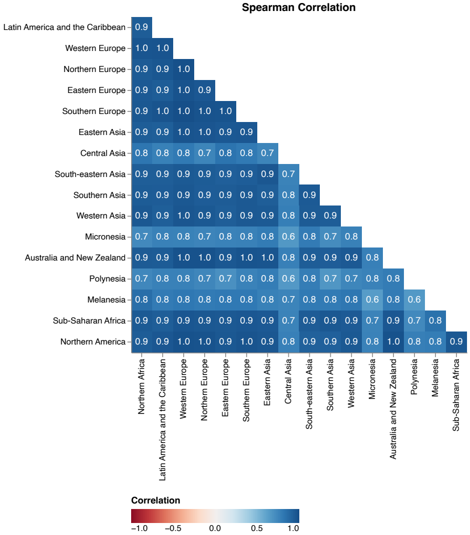
Notes: None of the original sentences contained emojis.

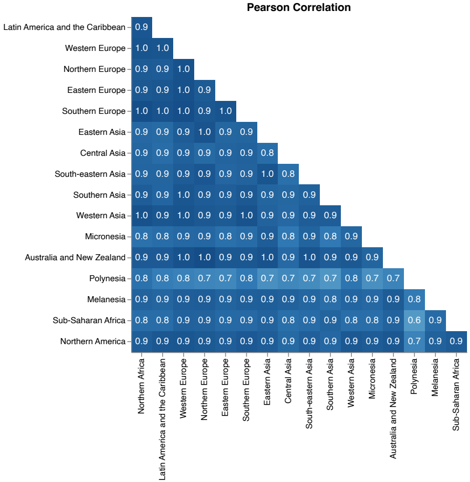
Geographical origins of geo-coded tweets



Our dataset covers tweets from all urban (density>4,000 people/km2) and 70% of suburban regions (density>200 people/km2)

Correlation of emoji frequency between different countries





Stage II: Regressor

Aim Reduce *predicted* emojis to basic emotions

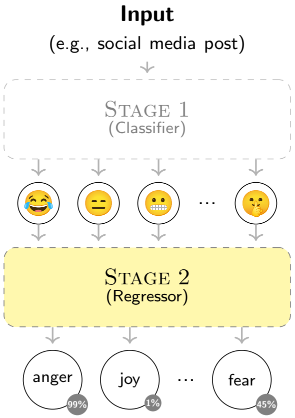
Approach Train a regressor to *predict affective intensities* based on *predicted* emojis

Data NRC lexicon as training set (Mohammad and Turney, 2013)

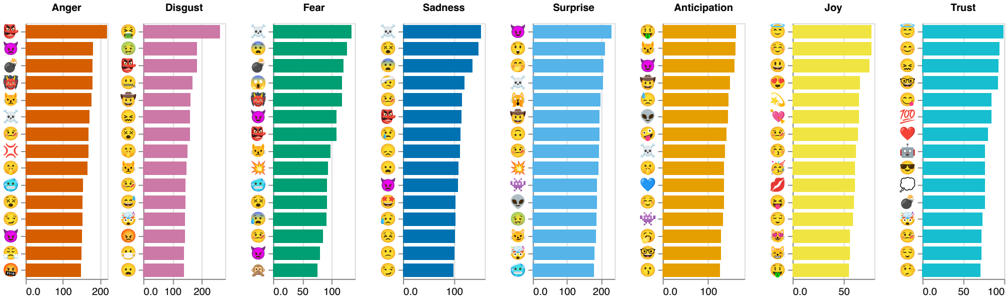
- Covers Plutchik's eight basic emotions (i.e., *sadness*, *anger*)
- Provides emotion scores for 5,975 English words

Model Gradient boosting

- Accounts for bounded nature of intensities
- Tree-based methods work well on related tasks (Felbermayr and Nanopoulos, 2016; Hartmann et al., 2019)



Emoji importances

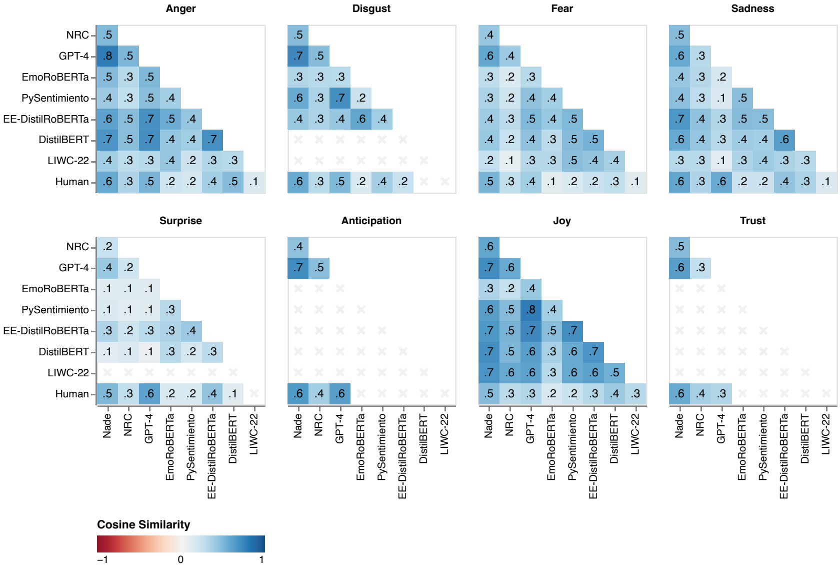


Validation - Match with human ratings

Metric	Method	Anger	Fear	Joy	Sadness	Surprise	Disgust	Trust	Anticipation
RMSE	NADE	**.33**	**.19**	**.32**	**.30**	**.31**	**.31**	**.22**	**.23**
	NRC	.36	.23	.35	.35	.32	.35	.27	.29
	LIWC-22	.38	.22	.35	.37	/	/	/	/
	EmoRoBERTa	.40	.24	.38	.38	.34	.37	/	/
	PySentimiento	.40	.25	.42	.39	.34	.43	/	/
	EE-DistilRoBERTa	.41	.27	.43	.40	.33	.36	/	/
	GPT-4	.48	.28	.43	.33	.35	.42	.29	.34
	DistilBERT	.48	.27	.52	.43	.34	/	/	/
	Human raters (SD)	.35	.26	.35	.34	.31	.34	.28	.31
MAE	NADE	**.24**	**.14**	**.23**	**.23**	**.24**	**.23**	**.16**	**.17**
	NRC	.27	.16	.24	.27	.26	.26	.19	.22
	LIWC-22	.28	.16	.23	.28	/	/	/	/
	EmoRoBERTa	.30	.16	.25	.29	.27	.27	/	/
	PySentimiento	.30	.17	.29	.30	.27	.32	/	/
	EE-DistilRoBERTa	.31	.18	.30	.31	.26	.26	/	/
	GPT-4	.38	.21	.32	.25	.28	.34	.21	.27
	DistilBERT	.38	.18	.40	.33	.28	/	/	/
	Human raters (mean abs dev)	.30	.16	.25	.29	.28	.29	.19	.25
cos θ	NADE	**.56**	**.51**	**.47**	**.57**	.49	**.56**	**.55**	**.64**
	NRC	.33	.26	.28	.33	.26	.29	.36	.37
	LIWC-22	.14	.11	.30	.14	/	/	/	/
	EmoRoBERTa	.21	.08	.16	.22	.16	.15	/	/
	PySentimiento	.19	.15	.23	.22	.17	.37	/	/
	EE-DistilRoBERTa	.37	.23	.30	.43	.43	.24	/	/
	GPT-4	.51	.38	.32	.55	**.59**	.49	.32	.62
	DistilBERT	.47	.26	.40	.34	.11	/	/	/

Validation - Match with human ratings

Validation - Comparison with scores from other converters



Advantages of Nade

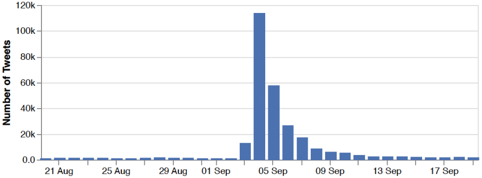
- **Implicit Human Ratings Through Emojis**
  - Utilizes emojis as implicit human ratings, available in large volumes online, reducing the need for human-annotated datasets.
- **Efficiency and Performance**
  - Offers a smaller and faster model compared to recent LLMs while achieving better predictive performance.
- **Overcoming Lexicon Limitations**
  - Emojifying language allows overcoming limited vocabularies of lexicon-based converters, making it effective for analyzing informal social media texts, including typos and slang.
- **Accessibility and Ease of Use**
  - Freely available via an easy-to-use app, requiring no programming knowledge, thus democratizing emotion analysis.
- **Validation and Benchmarking**
  - Validation through human ratings and benchmarked against state-of-the-art converters.

Study I: Brand perception tracking

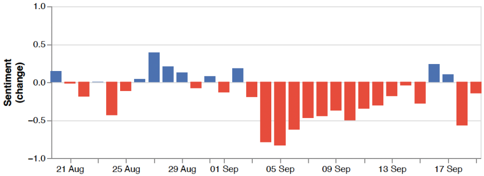
Managerial Problem: How do specific events affect brand perceptions?

- NADE allows to capture the dynamics of a diverse set of emotions
- Colin Kaepernick refused to stand during the national anthem in protest of wrongdoings against African Americans
- Nike picked him for 'Just do it' ad

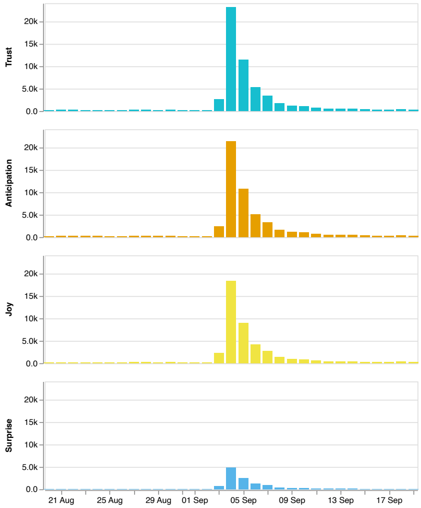
(a) Number of Nike Tweets per Day



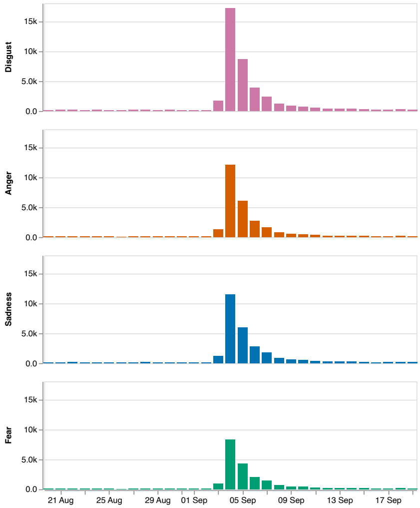
(b) Change in Valence



Study I: Brand perception tracking



(a) Positive Emotions



(b) Negative Emotions

Study II: Using emotions in online reviews to forecast demand

Managerial problem: Due to reputation inflation star-ratings tend to be overly positive ('positivity problem'; Rocklage, Rucker, and Nordgren (2021)); do reviews exhibit information that we can leverage?

- NADE allows to study fine grained differences in expressed emotional intensities
- 233 million Amazon reviews grouped by their star rating
- Emotional intensities vary predictably across the star ratings (e.g., greater anger for lower numeric ratings)
- Reviews with the same numeric ratings also vary considerably in the inferred emotions

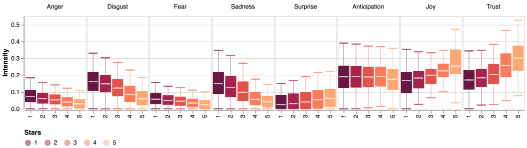


Figure: Emotional intensities extracted from Amazon reviews by star rating

Study II: Using emotions in online reviews to forecast demand

- Rocklage, Rucker, and Nordgren (2021) predict product success from the emotionality contained in online reviews.
- We build on their study and add measures of inferred emotions and emojis.
- Statistically significant improvements in predictive performance through the addition of Nade's scores (i.e., between feature sets 5 and 6; ΔVEcv=0.029, $t = 2.939$, $p < 0.05$). [6]

Feature Set	VEcv	Pearson Correlation (γ)	RMSE
1. Baseline	.116	.340	.619
2. Baseline + Nade (Emotion)	.149	.386	.607
3. Baseline + Nade (Emoji)	.209	.457	.585
4. Baseline + Nade (Emotion & Emoji)	.221	.471	.581
5. Baseline + text tools	.245	.495	.572
6. Baseline + text tools + Nade (Emotion & Emoji)	.274	.524	.561

[6] The included text tools are: Evaluative Lexicon, LIWC, and Vader. VEcv = variance explained in cross-validation, RMSE = root mean square error.

Study III: Creating emotionally engaging social media content

Managerial problem: What factors cause content to go viral?

- Literature consistently finds that while the high-arousal emotions (e.g., anger and anxiety) drive engagement, low-arousal emotions (e.g., sadness) have the opposite effect (Berger and Milkman, 2012; Berger, Moe, and Schweidel, 2023)
- We aim to replicate these findings by estimating the effects of emotions contained in video titles on the views of a video
- Sample: 1,330,531 videos by 3,966 channels
- Video titles are short! (m: 48, sd: 20)
- LIWC-22 can only label 32%, 32%, 31% of all video titles for anxiety, anger, and sadness (Nade: 95%, 93%, and 88%)

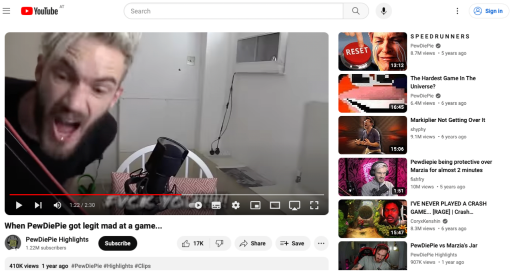


Figure: Example of an emotional YouTube video (and expressive title)

Study III: Creating emotionally engaging social media content

Managerial problem: What factors cause content to go viral?

- To assess the effect of emotions on video views, we estimate the following model:

$$
\begin{aligned}
ln(Views_i) = & \sum_j \beta_j \mathrm{NADE}_i^e \\
& + \sum_j \gamma_j VideoControls_i \\
& + \sum_j \delta_j TextControls_i \\
& + \sum_j \mu_j ChannelMonth_i + \epsilon_i,
\end{aligned}
$$

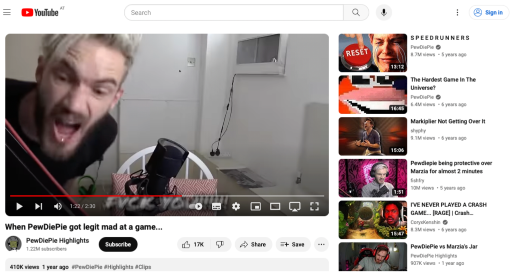


Figure: Example of an emotional YouTube video (and expressive title)

Study III: Creating emotionally engaging social media content

	ln(Views)	
	Model 1	Model 2
Anger (β_1)		.017*** (.002)
Anxiety (β_2)		.014*** (.001)
Sadness (β_3)		-.002 (.002)
Disgust (β_4)		.002 (.002)
Fear (β_5)		.032*** (.002)
Joy (β_6)		-.023*** (.002)
Anticipation (β_7)		-.024*** (.002)
Surprise (β_8)		-.005** (.002)
Trust (β_9)		-.022*** (.002)
PositiveTone (δ_1)	-.003 (.002)	
NegativeTone (δ_2)	.019*** (.002)	
VideoControls (γ)	✓	✓
ChannelMonth FE	✓	✓
TextControls	✓	✓
Number of videos	1,330,531	1,330,531
Number of channels	3,966	3,966
R^2	.730	.731

Note. Independent variables are standardized. Standard errors clustered by channel-month. FE = fixed effects. ***$p < .001$, **$p < .01$, *$p < .05$.

Study IV: Collective emotions

Managerial Problem: How do certain events manifest on social media?

- NADE facilitates massive scale affect detection
- Extract all eight emotions from all 575 million Reddit submissions
- Increasing levels of joy and decreasing levels of fear around Christmas
- Around Halloween, there is an increase in the levels of fear
- COVID-19 pandemic is manifested in the trajectories of various emotions (e.g., higher levels of sadness and fear)

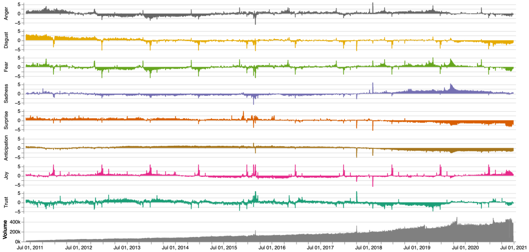
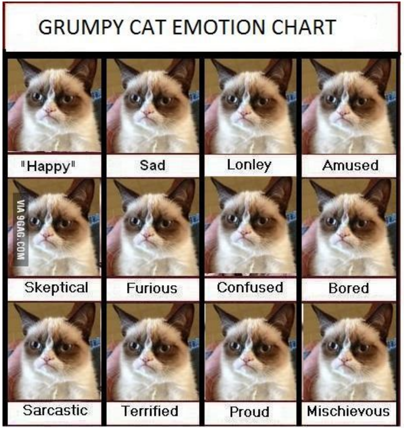


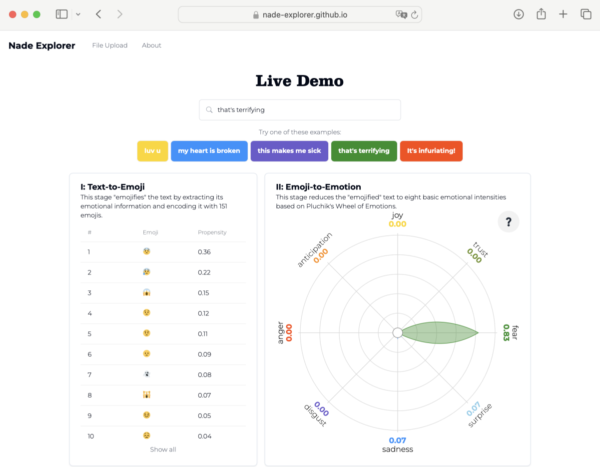
Figure: Emotions & volume of Reddit submissions over time

Summary

- Aids researchers & practitioners in *detecting* and *quantifying* emotional intensities from text
- Combines the *predictive performance* of LLMs and the *computational performance* of dictionaries
- We provide a web-app and packages for R and Python
- Is easily *extensible* to cover new languages, other emotional theories (e.g., VAD) or more emotions (e.g., relief)



Try it yourself!



Web-App: https://nade-explorer.inkrement.ai
Python package: https://github.com/inkrement/nade
R package: https://github.com/inkrement/nadeR

Natural Affect Detection (Nade)

Journal of Marketing

AMERICAN MARKETING ASSOCIATION

Impact Factor: 11.5
5-Year Impact Factor: 15.0

Journal Homepage

Available access | Research article | First published online January 13, 2025

EXPRESS: Natural Affect DEtection (NADE): Using Emojis to Infer Emotions from Text

Christian Hotz-Behofsits, Nils Wlömert, and Nadia Abou Nabout View all authors and affiliations

Accepted Manuscripts | https://doi.org/10.1177/00222429251315088

Contents | PDF/EPUB

Abstract

Emotions are central to consumer communications, and extracting them from user-generated online content is crucial for marketers, considering that such consumer opinions significantly shape brand perceptions, influence purchase decisions, and provide essential insights for marketing analytics. To leverage vast user-generated data, marketers and researchers require advanced text-to-emotion converters. However, existing tools for fine-grained emotion extraction face several limitations: Lexica are constrained by their dictionaries, machine learning models by human-annotated training data, and large language models by insufficient validation. As a result, marketing research still tends to rely on mere sentiment detection instead of extracting more nuanced emotions from text. This paper introduces Nade (Natural Affect DEtection), a novel text-to-emoji-to-emotion converter that first “emojifies” language and then converts these emojis into intensity measures of well-established, theory-grounded emotions. This approach addresses the limitations of existing tools by leveraging the inherent emotional information in emojis. Using human raters and state-of-the-art converters as benchmarks, the authors establish the benefits of exploiting emojis, validate Nade, and demonstrate its use in several marketing applications using data from various social media platforms. Users can apply the proposed converter through an easy-to-use online app and programming packages for Python and R.

https://journals.sagepub.com/doi/10.1177/00222429251315088

References I

Berger, Jonah and Katherine L. Milkman (2012). “What Makes Online Content Viral?” In: *Journal of Marketing Research* 49.2, pp. 192–205.

Berger, Jonah, Wendy W. Moe, and David A Schweidel (2023). “What Holds Attention? Linguistic Drivers of Engagement”. In: *Journal of Marketing* 87.5, pp. 793–809.

Bojanowski, Piotr et al. (2017). “Enriching Word Vectors with Subword Information”. In: *Transactions of the Association for Computational Linguistics* 5, pp. 135–146.

Ekman, Paul (1992). “An Argument for Basic Emotions”. In: *Cognition and Emotion* 6.3-4, pp. 169–200.

Felbermayr, Armin and Alexandros Nanopoulos (2016). “The Role of Emotions for the Perceived Usefulness in Online Customer Reviews”. In: *Journal of Interactive Marketing* 36.1, pp. 60–76.

Hartmann, Jochen et al. (2019). “Comparing Automated Text Classification Methods”. In: *International Journal of Research in Marketing* 36.1, pp. 20–38.

Joulin, Armand et al. (2016). “Bag of tricks for efficient text classification”. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 427–431.

Kranzbühler, Anne-Madeleine et al. (2020). “Beyond Valence: A Meta-Analysis of Discrete Emotions in Firm-Customer Encounters”. In: *Journal of the Academy of Marketing Science* 48.3, pp. 478–498.

Mohammad, Saif M. and Peter D. Turney (2013). “Crowdsourcing a Word-Emotion Association Lexicon”. In: *Computational Intelligence* 29.3, pp. 436–465.

Plutchik, Robert (1980). “Theories of Emotion”. In: Elsevier. Chap. A General Psychoevolutionary Theory of Emotion, pp. 3–33.

Rocklage, Matthew D., Derek D. Rucker, and Loran F. Nordgren (2021). “Mass-Scale Emotionality Reveals Human Behaviour and Marketplace Success”. In: *Nature Human Behaviour* 5.10, pp. 1–7.

References II

Russell, James A. (2003). "Core affect and the psychological construction of emotion". In: *Psych. Rev.* 110.1, p. 145.

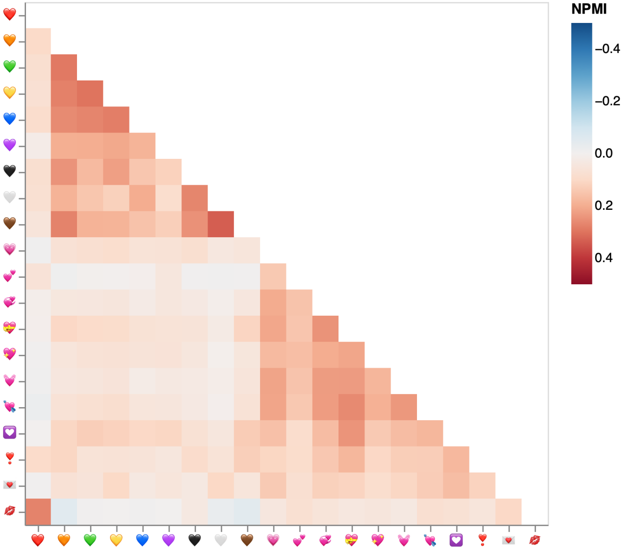
Vosoughi, Soroush, Deb Roy, and Sinan Aral (2018). "The Spread of True and False News Online". In: *Science* 359.6380, pp. 1146–1151.

Yin, Dezhi, Samuel D. Bond, and Han Zhang (2014). "Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews". In: *MIS Quarterly* 38.2, pp. 539–560.

Understanding Emotions and Affect

- **Affect**
  - *Definition:* An encompassing term for emotional experiences, including emotions and moods.
  - *Characteristics:* Represents the broader spectrum of feeling states.
- **Core Affect**
  - *Definition:* Fundamental, non-reflective state characterized by valence and arousal.
  - *Characteristics:* Persistent, background emotional state not directed towards anything specific.
- **Moods**
  - *Definition:* Prolonged emotional states that are less intense and less specific than emotions.
  - *Characteristics:* Can last for hours or days, influencing overall feeling and perception.
- **Emotions**
  - *Definition:* Mental states arising from core affect changes, influenced by cognitive appraisals.
  - *Characteristics:* Specific, often intense, and typically directed at a particular object.

Why not use emojis directly?



Correlations between Nade and Sentiment Scores