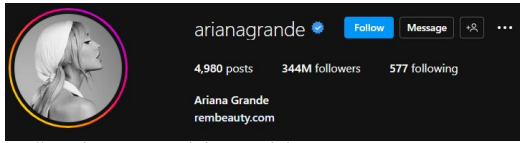
A Computational Analysis Of Advertising Language On Social Media

Lennart Rösemeier ▪ l.roesemeier@gmail.com ▪ +49 15751089270

What Is An Influencer?

- Creates digital content that captures the attention of thousands of viewers (Zulli, 2018)
- Possibility to monetize their content
- Independent from specific gatekeepers like TV, movie, or music industry (Hearn and Schoenhoff, 2015)
- Depending on the amount of followers, they can be subdivided in (Ruiz-Gómez, 2019):
  - Micro influencers (5,000 - 100,000 followers)
  - Power middle influencers (100,001 - 500,000 followers)
  - Macro influencers (500,001 - 1,000,000)
  - Mega influencers (over 1,000,000 followers)

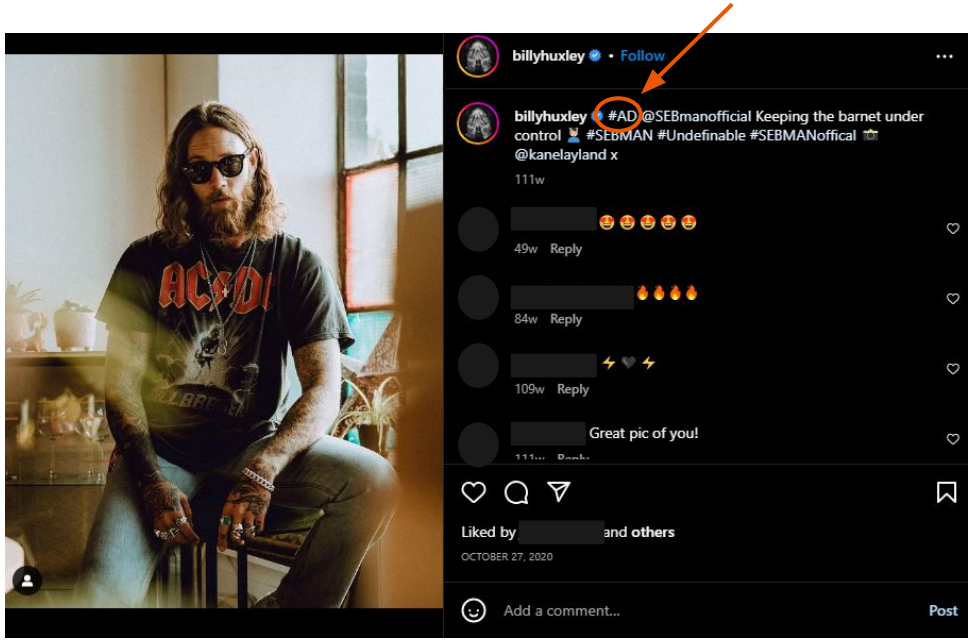
https://www.instagram.com/billyhuxley/



https://www.instagram.com/arianagrande/

What Is Advertised Content On Social Media?

- Integrating ads in digital content (de Gregorio & Goatana, 2021):
  - Sponsored advertising
  - Product sampling
  - Affiliate marketing
- Integrated ads has to be disclosed via, *e.g.*, the signs “Ad”, “Paid Sponsorship”, or similar (de Gregori & Goanta, 2021; Kim et al., 2021)



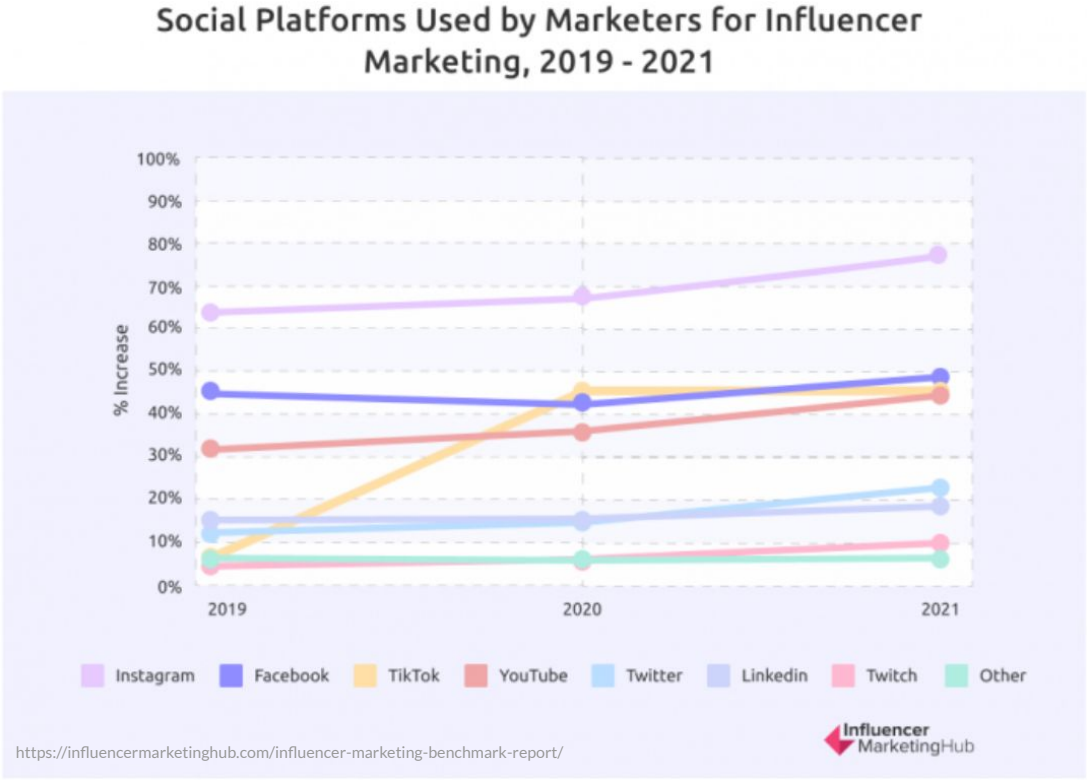
https://www.instagram.com/p/CG2gJjkhbbZ/

What Is Undisclosed Advertisement On Social Media?

- Social media marketing can only become effective when influencers are trusted sources of information (Kim et al., 2021)
- Authenticity is one of the key role for influencers to sustain long term recognition (Khamis et al., 2017)
- Thus, influencers often do not disclose their advertisements

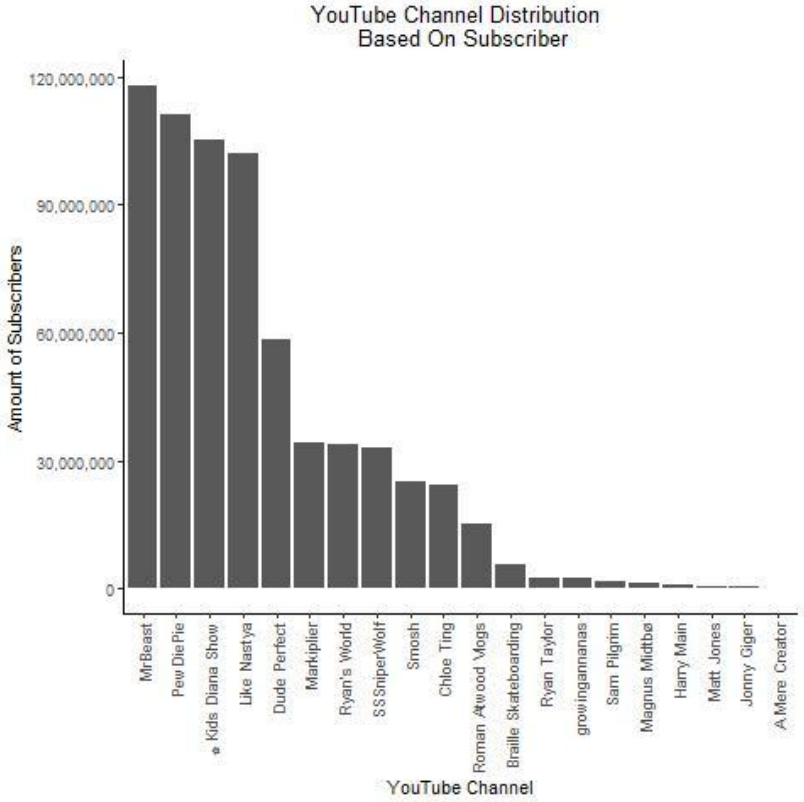
https://www.instagram.com/p/ClrAZdMI2T0/

Data Collection



Data Collection

- Selection of YouTube channels which hosting language is in English
- Selected the biggest YouTube channels + minor mega and major macro channels (subscriber amount range: 225k - 118M)
- Categories: gaming, vlogs, toys, extreme sport, fitness, crafting



Data Collection

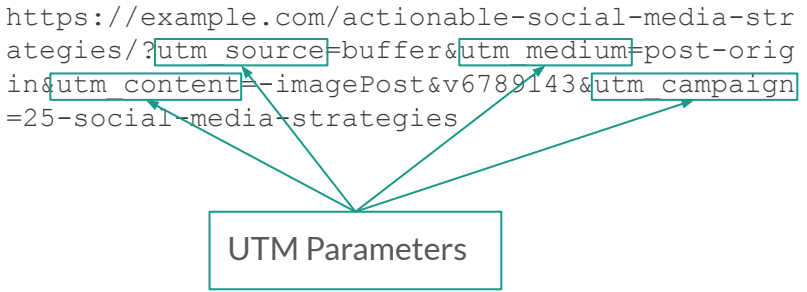
- Newest 50 videos of each channel due to limited time resources
  - ~1h to transcribe 1000 videos
- Variables of main interest:
  - video description
  - subtitles
  - video ID
- Final dataset: 894 videos and 6 variables
- Data cleansing

Video ID	Description	Publishing Date	Channel Name	Channel ID	Subtitles
I4AgeDIrHGY	“bucket list us back bungee jumping feeding elephants paragliding chilling with penguins and rhino rescue in south africa has it all [...]”	2021-05-17T21:59:59Z	Dude Perfect	UCRijo3ddMTht_IHyNSNXpNQ	“i don’t think i could this. ok. i’m going. [crowd cheering][]music playing] ladies and gentleman, as you can see on my phone we are on our way [...]”

Data Collection

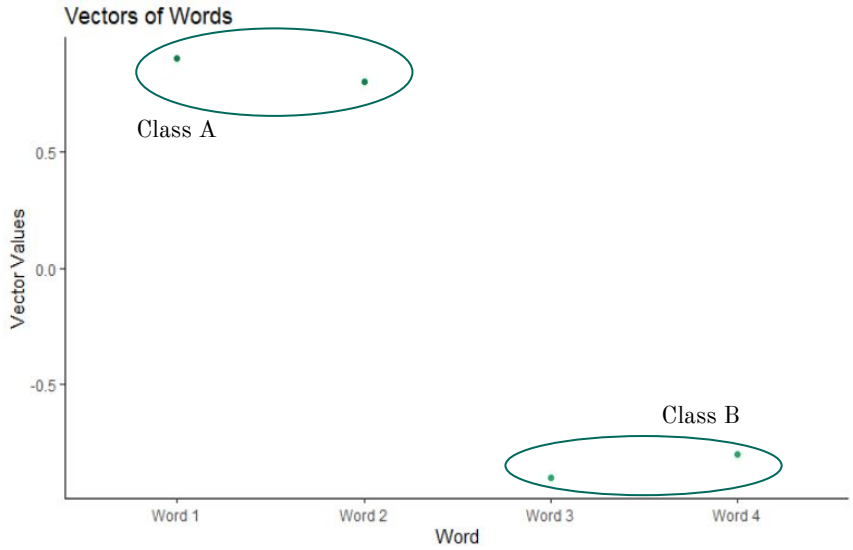
- Training dataset was created partially manually
- Detection of disclosed ads, however, is automatized by specific characters and Urchin Tracking Module (UTM) patterns at URL links
- Training data consist of 151 videos with advertisements
  - 36 disclosed
  - 115 undisclosed

```
ads <- top_videos %>%
  filter(grepl('(^|\\s)ad($|\\s)|advertisment|sponsorship
                |sponsored|sponsoring|affiliate',
                description))
```



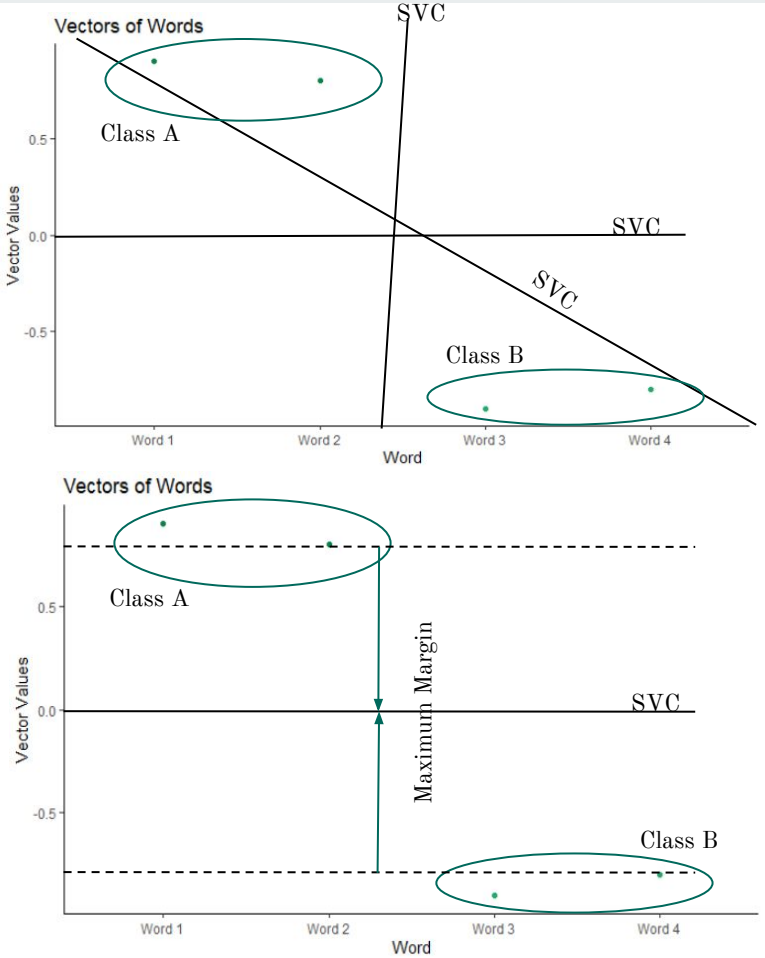
Analysis

- Support Vector Classifier works best for YouTube advertisement analysis (Swart et al., 2020)
- Vector values get allocated to words
- Similar words get mapped to similar points within a geometrical space
- Points get clustered
- Each cluster get assigned to a class



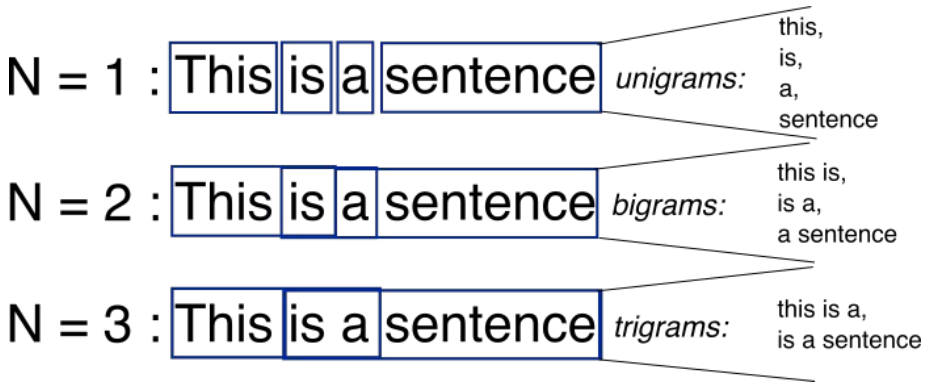
Analysis

- Classes are divided by a hyperplane
  - Hyperplane defines also the boundaries of the classes
  - Mostly, more than one option for a hyperplane
- Right hyperplane is chosen by maximum margin of vectors to the hyperplane
- Chosen hyperplane = Support Vector Classifier



Analysis

- Implementing other NLP features like bag-of-words or n-grams can improve algorithm performance (Marafino et al., 2014; Swart et al., 2020)
- To detect hidden advertisements the order of words in a sentence is important
  - “get exclusive deal”
  - “sponsored by”
- N-grams respect the word order to predict the next word in a sentence
- Thus, n-gram may improve the performance of the algorithm



https://stackoverflow.com/questions/18193253/what-exactly-is-an-n-gram

Results

- Performance measurements
  - Precision: $\frac{TP}{TP + FP}$, (true positives)
  - Recall: $\frac{TP}{TP + FN}$, (true positives over all positive cases)
  - F-Score: $2 \times \frac{precision \times recall}{precision + recall}$, (harmonic mean of precision and recall)

	Precision	Recall	F-Score
SVM	87%	94%	90%
SVM w/ Bigram Feature	94%	86%	90%

- The bigram feature does not improve the SVM algorithm, judged by the F-Score

Results

- SVM detects 315 videos with hidden advertisements
- Whether in the video description or subtitles, in some cases both
- Approximately 35 % of videos contain undisclosed advertisements

Video ID	Description	Publishing Date	Channel Name	Channel ID	Subtitles
RSHcpUvVHyY	“[...] get a 1 year supply of vitamin D + 5 individual travel packs free with your first purchase, got to https://athleticgreens.com/mattjones [...]”	2022-07-17T17:00:17Z	Matt Jones	UCXEuQ_O6BUWwhVhcMyXNHgg	“oh it’s spiky welcome back to the playground this epic ongoing build project in the last video you probably saw this hip lurking in the background [...]”

Discussion

- SVM is a reliable algorithm to detect hidden advertisements on YouTube (F-Score: 90 %, ~35 % of videos contain undisclosed advertisements)
- Fully automatized influencer detection in social media networks
- Image & video analysis
- Subtitles are disabled or do not exist
- Validation (*e.g.*, training dataset, algorithm)

Main Challenges

- Scrape limitations by platforms (especially: Meta)
  - Writing new packages
  - Pricing
- Different data formats:
  - Text
  - Image
  - Video
- Available (training) datasets of advertising language for algorithms

Main Challenges

- Different content on each social media platform
- Different product categories
  - Different marketing strategies & language
- Dictionaries (slang, irony, etc.)
- Change of marketing strategies
  - Models can get unreliable

Top 10 Influncers 2022

YouTube Channels	Instagram Accounts
Mr. Beast	Instagram
PewDiePie	Cristiano Ronaldo
Kids Diana Show	Lionel Messi
Like Nastya	Kylie Jenner
Dude Perfect	Selena Gomez
Markiplier	Dwayne ‘The Rock’ Johnson
Rayn’s World	Ariana Grande
SSSniperWolf	Kim Kardashian
Smosh	Beyoncé
Chloe Ting	Kendall Jenner

https://influencermarketinghub.com/youtube-influencers/
https://influencermarketinghub.com/top-instagram-influencers/

References

- de Gregorio, G., & Goanta, C. (2022). The Influencer Republic: Monetizing Political Speech on Social Media. *German Law Journal*, *23*(2), 204-225.
- Hearn, A., and Schoenhoff, S. (2015). From celebrity to influencer; Tracing the diffusion of celebrity value across the data stream. *A companion to celebrity*, 194-212.
- Khamis, S., Ang, L. and Welling, R.,. (2017). Self-branding, 'micro-celebrity' and the rise of Social Media Influencers. *Celebrity Studies*, 8(2), 191-208
- Kim, S., Jiang, J. Y., & Wang, W. (2021). Discovering undisclosed paid partnership on social media via aspect-attentive sponsored post learning. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (pp. 319-327).
- Marafino, B. J., Davies, J. M., Bardach, N. S., Dean, M. L., and Dudley, R. A. (2014). N-gram support vector machines for scalable procedure and diagnosis classification, with applications to clinical free text data from the intensive care unit. *Journal of the American Medical Informatics Association*, 21(5):871–875.
- Ruiz-Gomez, A. (2019) Digital Fame and Fortune in the age of Social Media: A Classification of social media influencers *aDResearch ESIC*. Nº 19 Vol 19
- Swart, M., Lopez, Y., Mathur, A., & Chetty, M. (2020, April). Is this an ad?: Automatically disclosing online endorsements on youtube with adintuition. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).
- Zulli, D. (2018). Capitalizing on the look: insights into the glance, attention economy, and Instagram. *Critical Studies in Media Comunication*, 35(2), 137-150.