

Setting them up for failure – How customer expectations collide with economic realities of text analytics



About me



Patrick Bunk

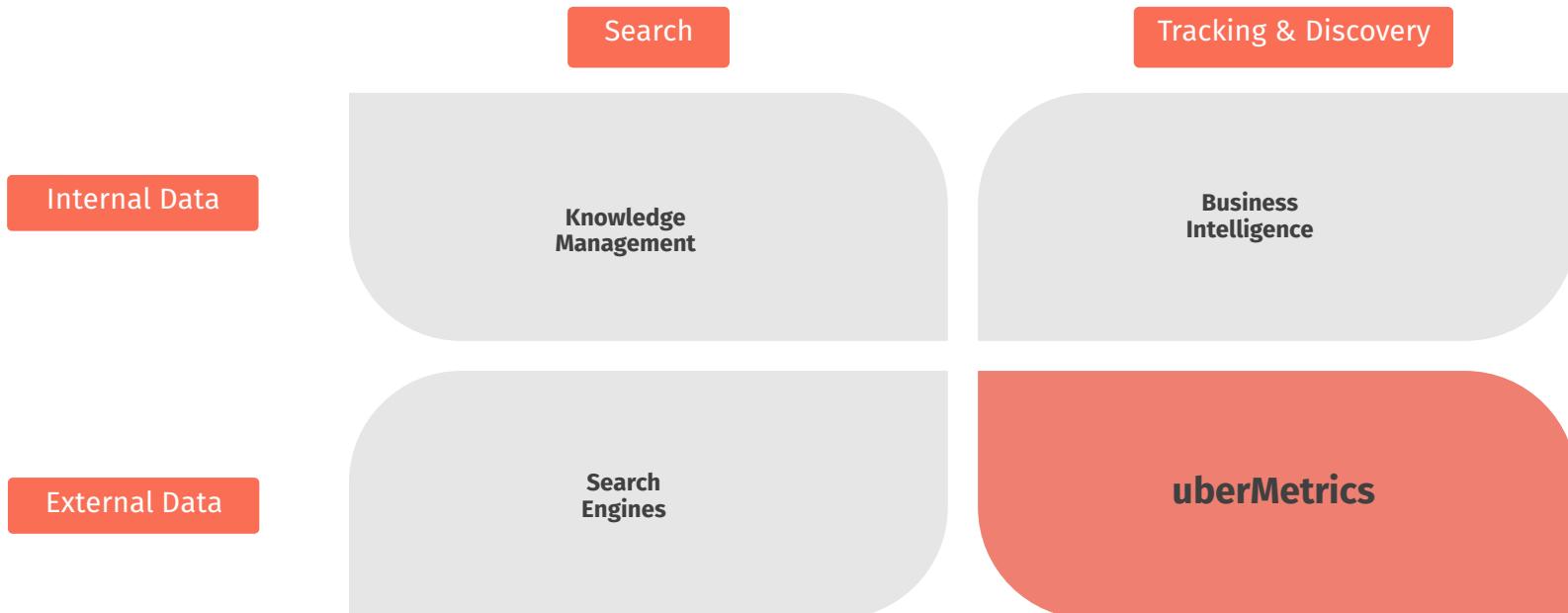
CEO
uberMetrics Technologies GmbH

 Economist

 Researcher DFG SFB649 Economic Risk

 Founder and CEO of uberMetrics

Long Term



The (small) Problem

- Companies want to know what customers & the public discuss
- Brands, Products & Companies

Clippings / Press Review modernized

- Customers – Analytics, Alerts
- Competition - Content Success, Benchmarking, Alerts
- Supply Chain – Alerts

How much data do they need?

- Mean 407,830 articles/month
- Median 26,928 articles/month

What filters do they have?



(Social) Media Monitoring

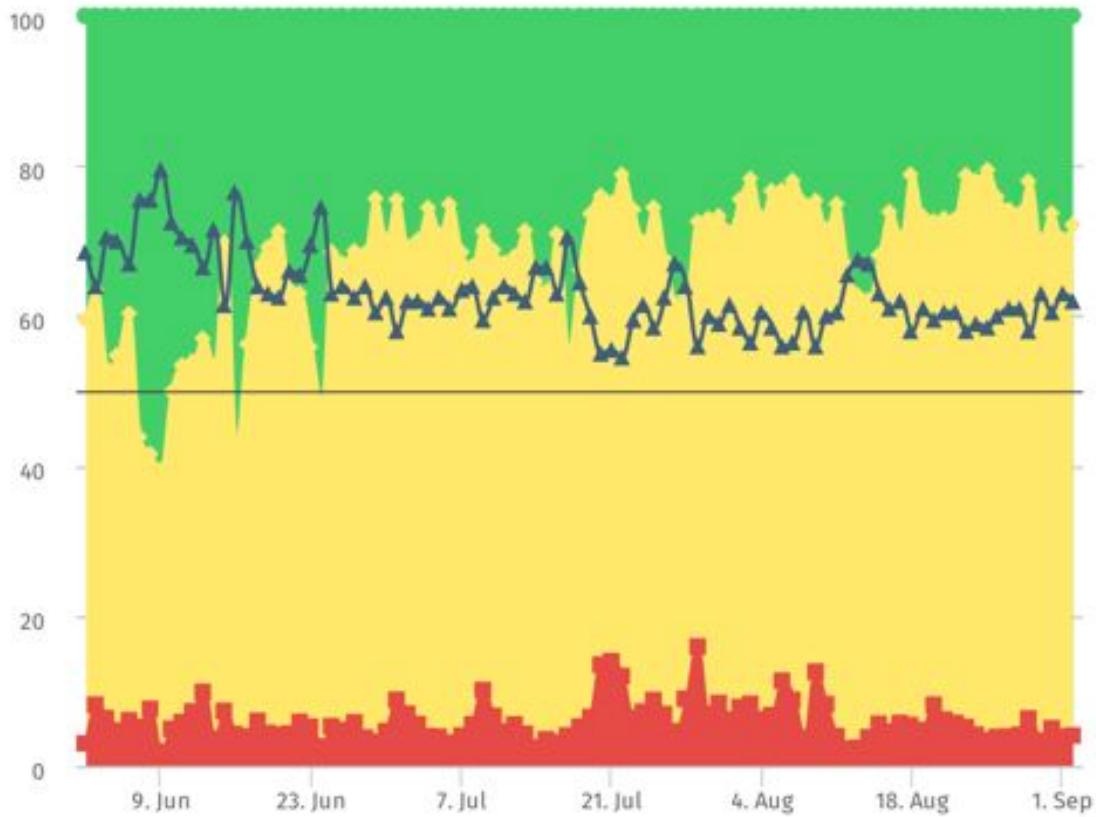
Counting mentions

- over time
 - by source
 - by segment
 - by author
 - benchmarked with competitors
 - virality
-
- topics
 - sentiment distribution



Sentiment

- simplifies complex realities intuitively
- meaningful categories
- summable
- expectations gap
- quality varying over time and domains
- Better technology to make customers happy?



Sentiment-Improvements

- standard sentiment models 70-80% (customer measured) for about 0€
- labor-based baseline for sentiment & topics 1€/article
- Tailor-made solution
 - Build a corpus, modelling
 - train a custom model
 - sell a proprietary classifier
 - Minimum 60k€ setup + 60k€ recurring for up to x articles/year
= one full time employee
- Crowd-based tagging 0.05€/article



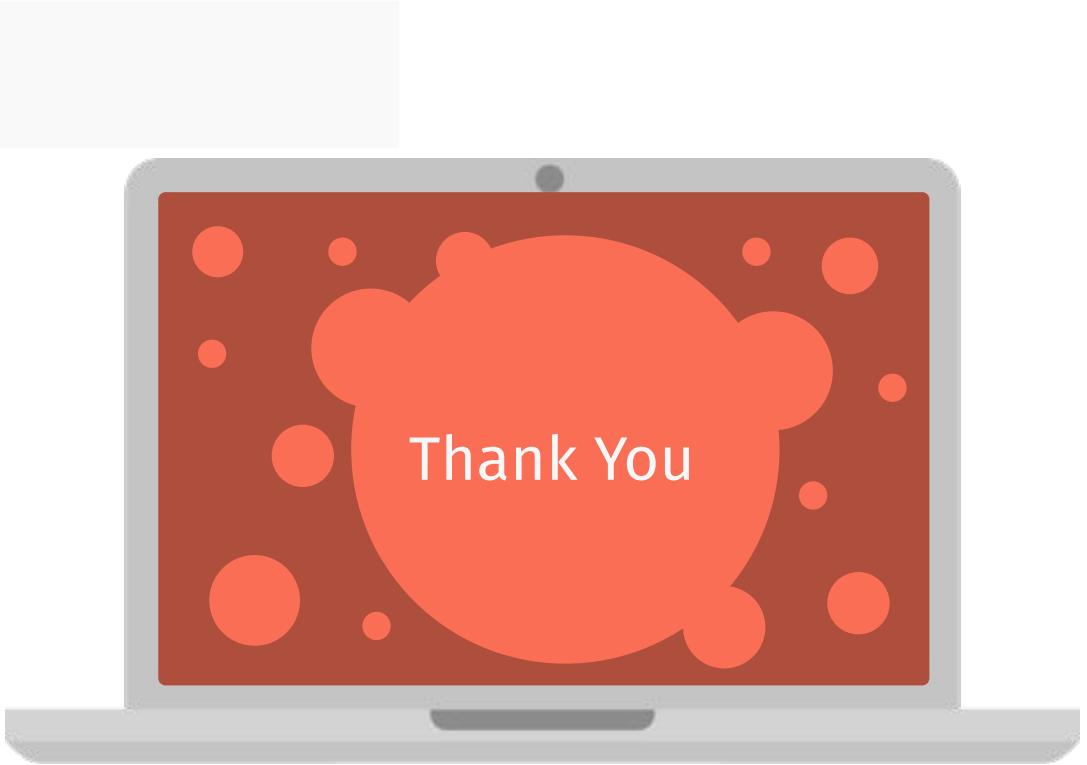
Sentiment-Economics

- fixed cost of creating algorithm
- fixed training-cost (time of an expert)
 - Limited supply → constant cost
- deflationary economics of processing power -> execution costs → 0
- Customer Acquisition Costs
 - complex product
 - contractual limits of ML
 - consulting more pre- than post- signing



Text-Mining Challenges

- Testing algorithms on humanities majors
- coping with failure gracefully
- focus on generalized solutions
- easy adaptability by the end-user
- be aware of manual labor substitute
- Tailor-made Mining is at a local maximum pre scalable product
- Automation through knowledge should be socially beneficial
- commercial domain public high-quality data (Trade Registers etc)
- copyrights and transferring data (Google Advantage)



/uberMetrics



/uberMetrics