

# Reviewing rapid prototype candidates

for data-driven projects

Sebastian Sauer

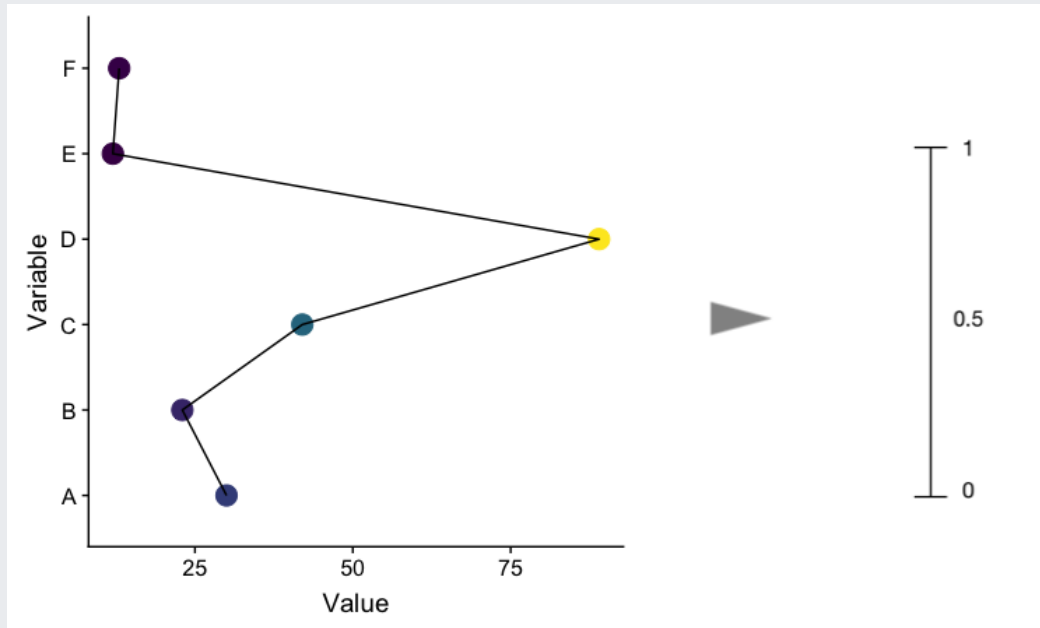
# Overview

1. **Employee retention:** Predict employee propensity to leave the company
2. **Predictive competition:** Compare the predictive performance of traditional/novel models
3. **Social Listening:** Quantify brand opinion (and related emotions)
4. **Objective organization climate:** Build text-based model for organization climate

1. Employee retention: Predict employee propensity to leave the company

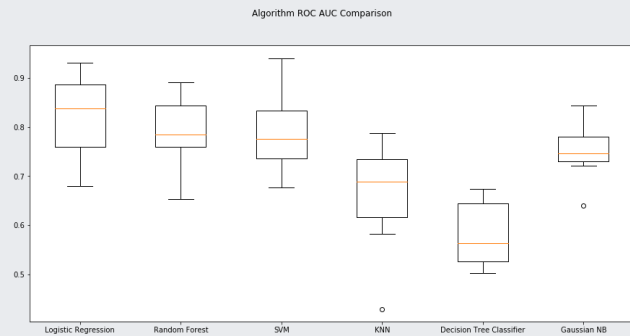
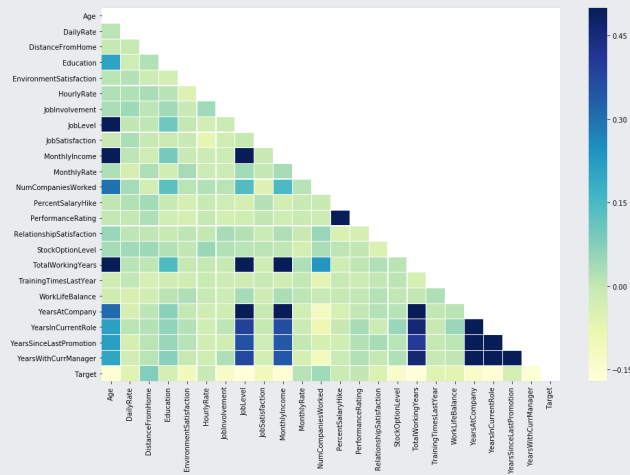
# Input: Employee's data, output: leave propensity

! data privacy



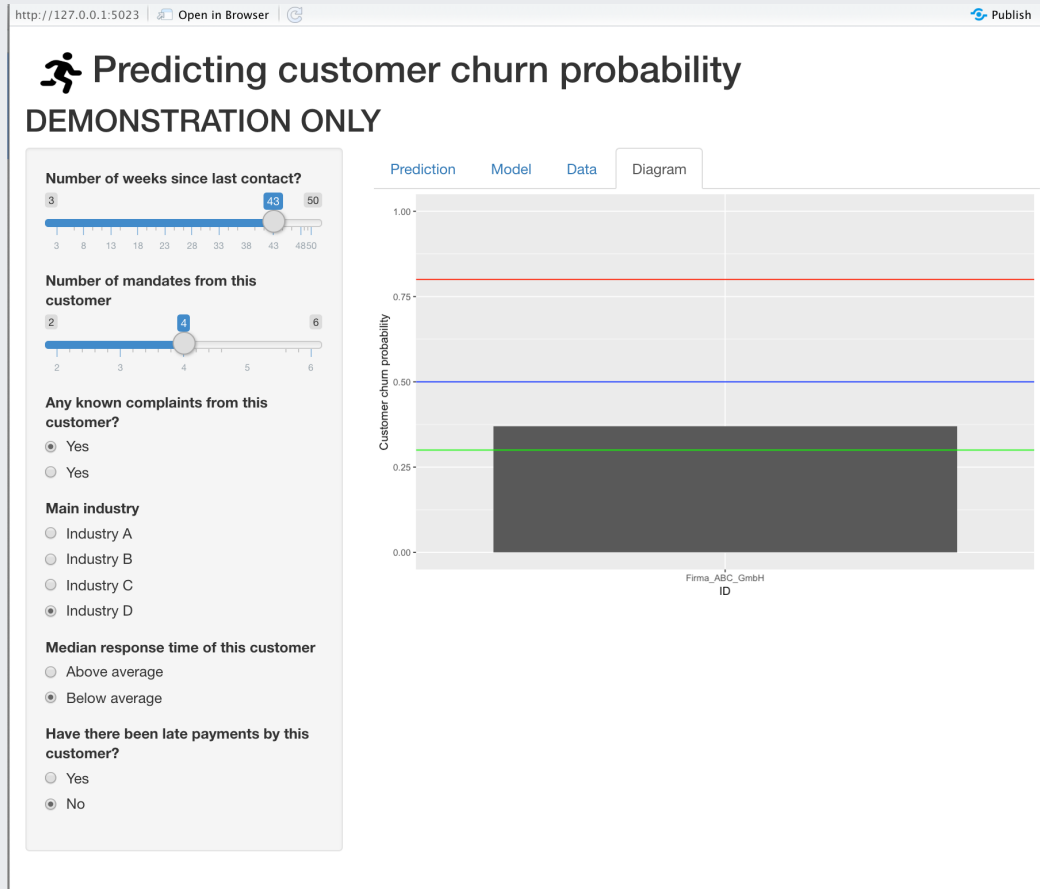
# Industry example: employee retention at IBM

IBM artificial intelligence can predict with 95% accuracy which workers are about to quit their jobs. See this [case study](#).



Source: [CNBC](#), [TowardsDataScience](#)

# See live app




2. Predictive competition:  
Compare the predictive  
performance of traditional/novel  
models

# Case study -- Predicting therapy success (1/2)

## ORIGINAL ARTICLE

### **Prediction of outcome in internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: A machine learning approach**

Fabian Lenhard<sup>1,2</sup>  | Sebastian Sauer<sup>3</sup> | Erik Andersson<sup>1</sup> | Kristoffer NT Månsson<sup>4,5</sup> | David Mataix-Cols<sup>1,2</sup> | Christian Rück<sup>1,2</sup> | Eva Serlachius<sup>1,2</sup>

Lenhard, F., Sauer, S., Andersson, E., Månsson, K. N., Mataix-Cols, D., Rück, C., & Serlachius, E. (2018). Prediction of outcome in internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: A machine learning approach. *International Journal of Methods in Psychiatric Research*, 27(1), e1576.  
<https://doi.org/10.1002/mpr.1576>

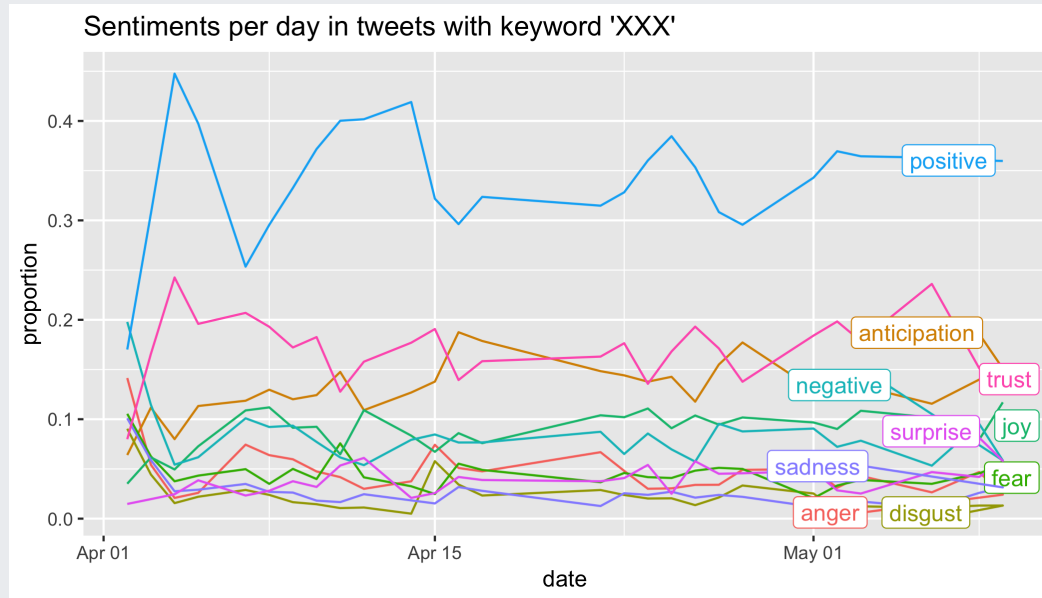


# Case study -- Predicting therapy success (2/2)

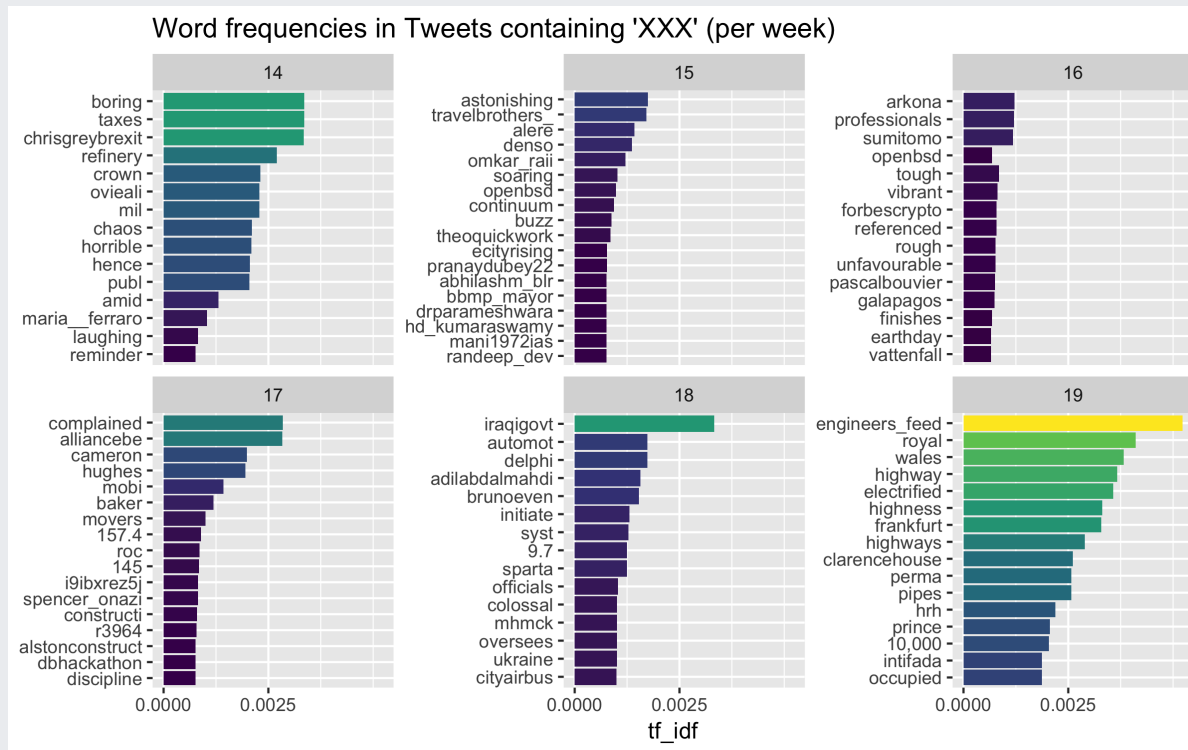


Social Listening: Quantify brand  
opinion (and related emotions)

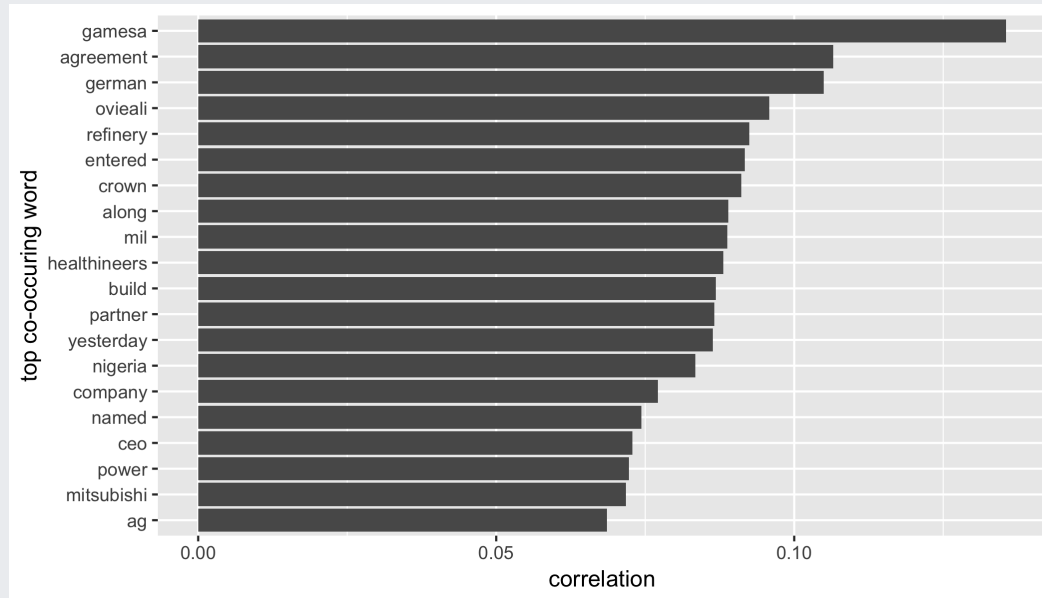
# Emotions in tweets with keyword 'XXX'



# Word frequencies in tweets containing 'XXX'



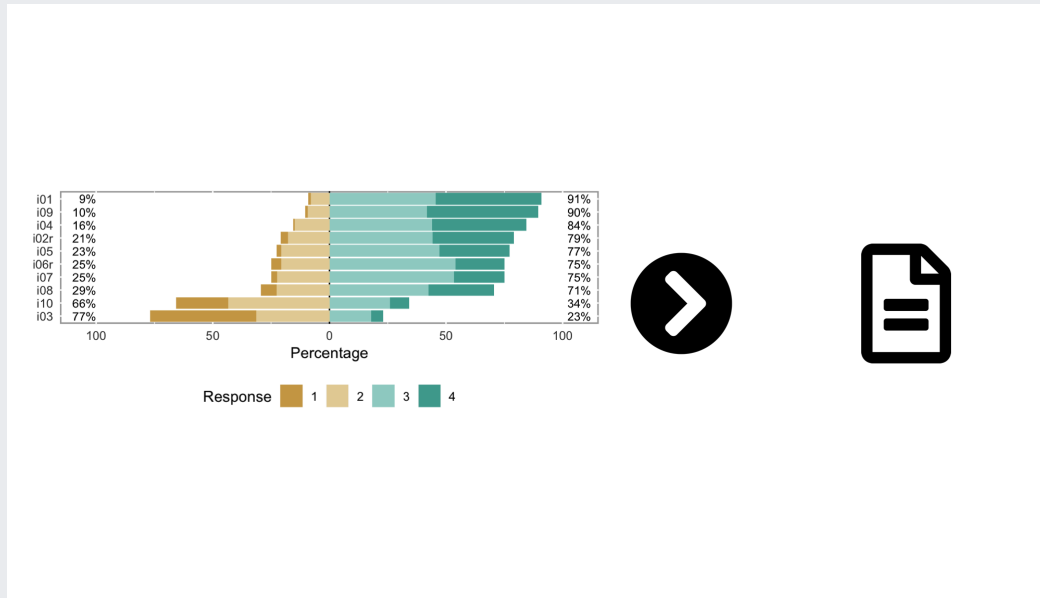
# Which words correlate with 'XXX' most strongly?



Phi correlation, per tweet, based on tweet data presented previously

4. Objective organization  
climate: Build text-based model  
for organization climate

# Calibrate words to measure organizational climate



Sebastian Sauer

 [sebastiansauer](#)

 <https://data-se.netlify.com/>

 [sebastian.sauer@data-divers.com](mailto:sebastian.sauer@data-divers.com)

 [sauer\\_sebastian](#)

CC-BY



# Reproducibility

- Versions of employed software as of 2019-05-29, running this OS: macOS Mojave 10.14.5.
- Built with R, R version 3.6.0 (2019-04-26), RStudio 1.2.1335, xaringan, on the shoulders of giants
- Icons are from [FontAwesome](#), licenced under CC-BY-4 ([details](#))
- R-Packages used: assertthat\_0.2.1, backports\_1.1.4, broom\_0.5.2, caret\_6.0-84, cellranger\_1.1.0, class\_7.3-15, cli\_1.1.0, codetools\_0.2-16, colorspace\_1.4-1, crayon\_1.3.4, data.table\_1.12.2, digest\_0.6.18, dplyr\_0.8.0.1, DT\_0.5, evaluate\_0.13, forcats\_0.4.0, foreach\_1.4.4, generics\_0.0.2, ggplot2\_3.1.1, glue\_1.3.1.9000, gower\_0.2.0, gridExtra\_2.3, gtable\_0.3.0, gtrendsR\_1.4.3, haven\_2.1.0, hms\_0.4.2, htmltools\_0.3.6, htmlwidgets\_1.3, httr\_1.4.0, icon\_0.1.0, ipred\_0.9-9, iterators\_1.0.10, jsonlite\_1.6, knitr\_1.22, labeling\_0.3, lattice\_0.20-38, lava\_1.6.5, lazyeval\_0.2.2, lubridate\_1.7.4, magrittr\_1.5, MASS\_7.3-51.4, Matrix\_1.2-17, ModelMetrics\_1.2.2, modelr\_0.1.4, munsell\_0.5.0, nlme\_3.1-139, nnet\_7.3-12, pillar\_1.3.1, pkgconfig\_2.0.2, plyr\_1.8.4, prodlim\_2018.04.18, purrr\_0.3.2, R6\_2.4.0, Rcpp\_1.0.1, readr\_1.3.1, readxl\_1.3.1, recipes\_0.1.5, reshape2\_1.4.3, rlang\_0.3.4, rmarkdown\_1.12.6, rpart\_4.1-15, rprojroot\_1.3-2, rstudioapi\_0.10, rvest\_0.3.3, scales\_1.0.0, sessioninfo\_1.1.1.9000, stringi\_1.4.3, stringr\_1.4.0, survival\_2.44-1.1, tibble\_2.1.1, tidyr\_0.8.3, tidyselect\_0.2.5, tidyverse\_1.2.1, timeDate\_3043.102, viridisLite\_0.3.0, withr\_2.1.2, xaringan\_0.9, xaringantheme\_0.2.0, xfun\_0.7, xml2\_1.2.0, yaml\_2.2.0
- Last update 2019-05-29