

Reviewing rapid prototype candidates

for data-driven projects

Sebastian Sauer

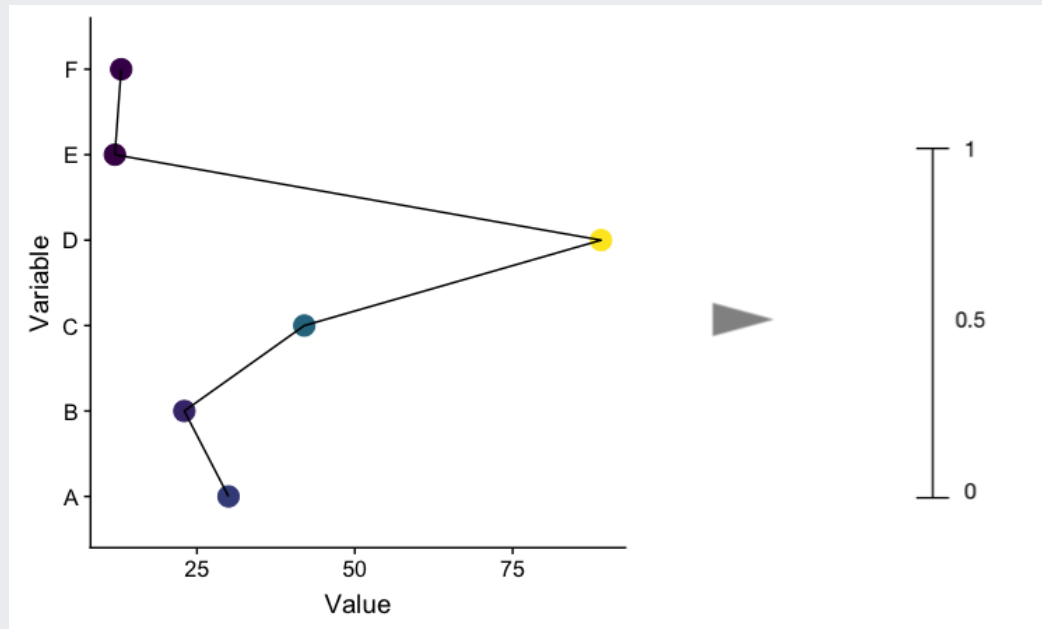
Overview

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

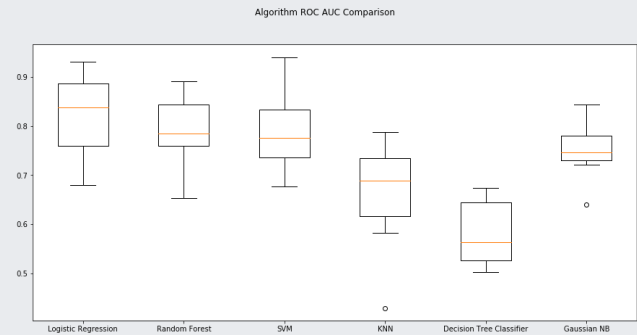
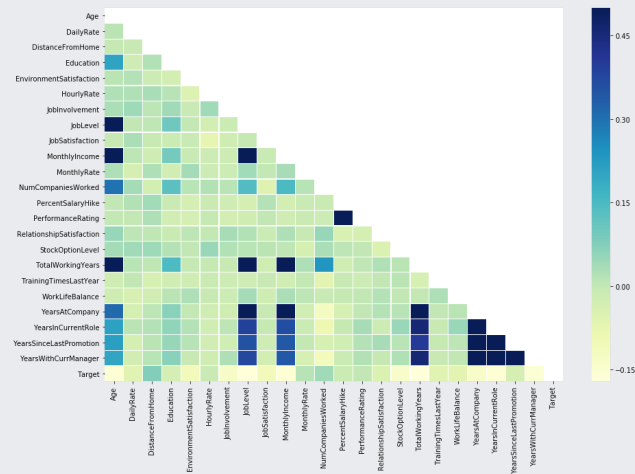
Input: Employee's data, output: leave propensity

! data privacy



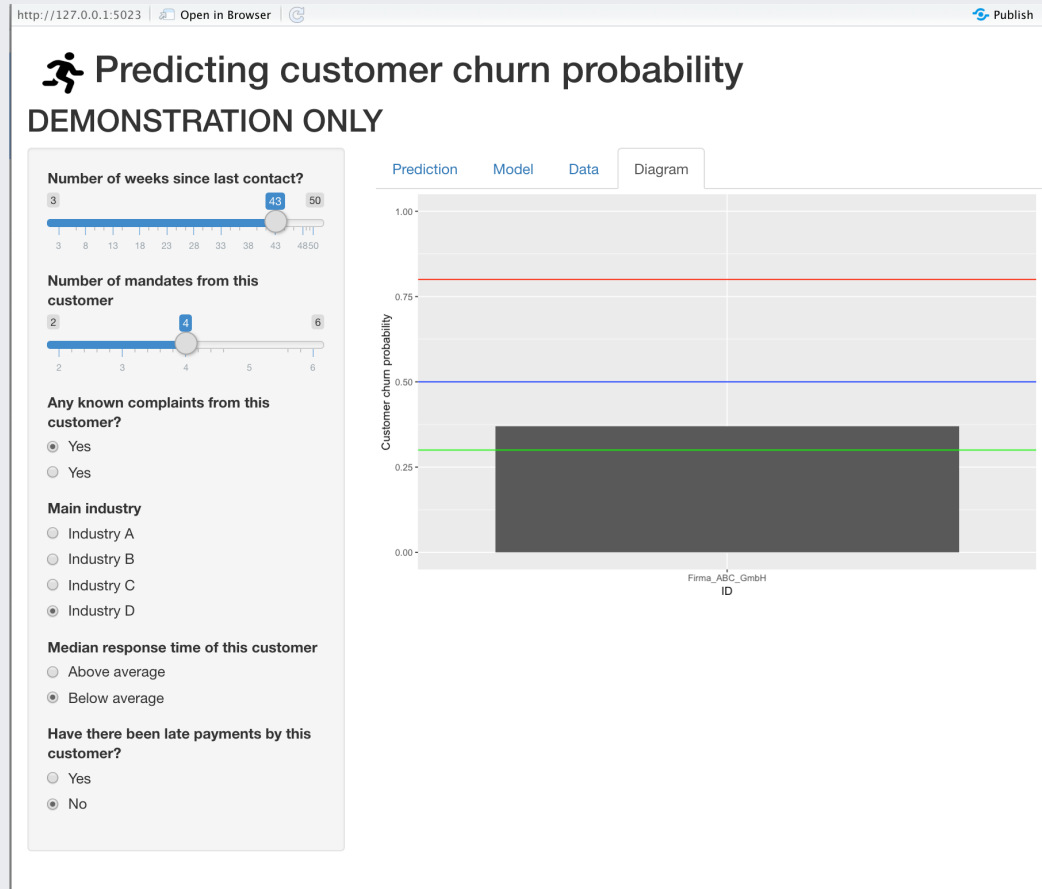
Industry example: employee retention at IBM

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)

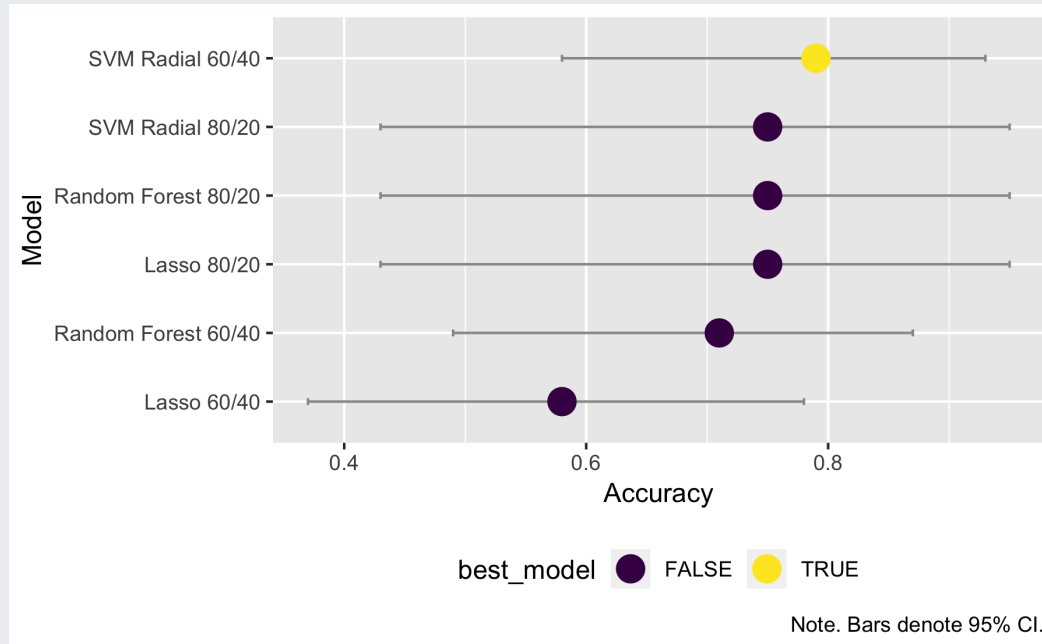
ORIGINAL ARTICLE

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

Fabian Lenhard^{1,2}  | Sebastian Sauer³ | Erik Andersson¹ | Kristoffer NT Månsson^{4,5} | David Mataix-Cols^{1,2} | Christian Rück^{1,2} | Eva Serlachius^{1,2}

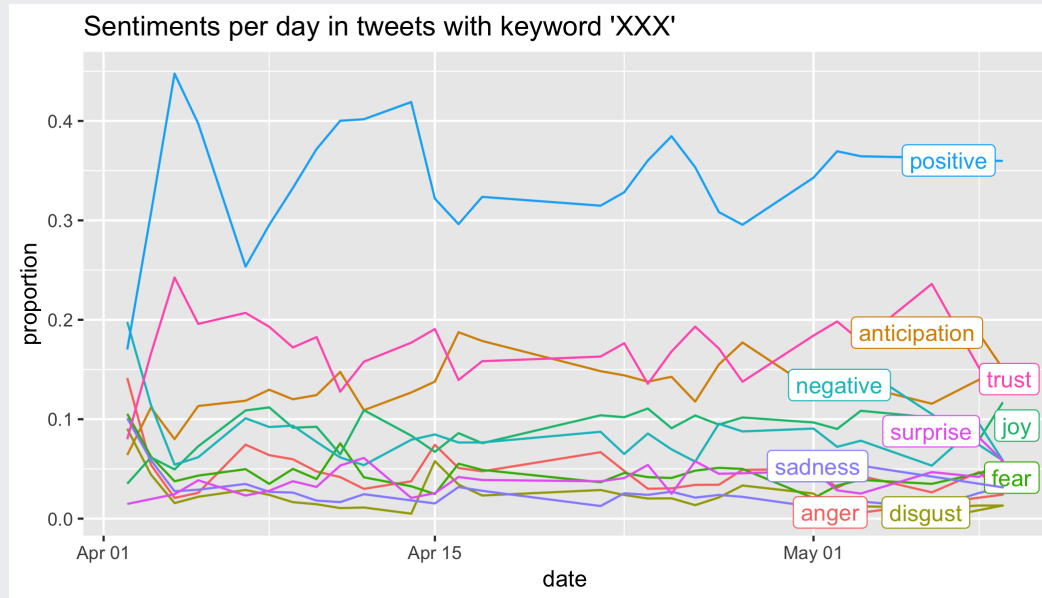
Lenhard, F., Sauer, S., Andersson, E., Månsson, K. N., Mataix-Cols, D., Rück, C., & Serlachius, E. (2018). Prediction of outcome in internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: A machine learning approach. *International Journal of Methods in Psychiatric 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)

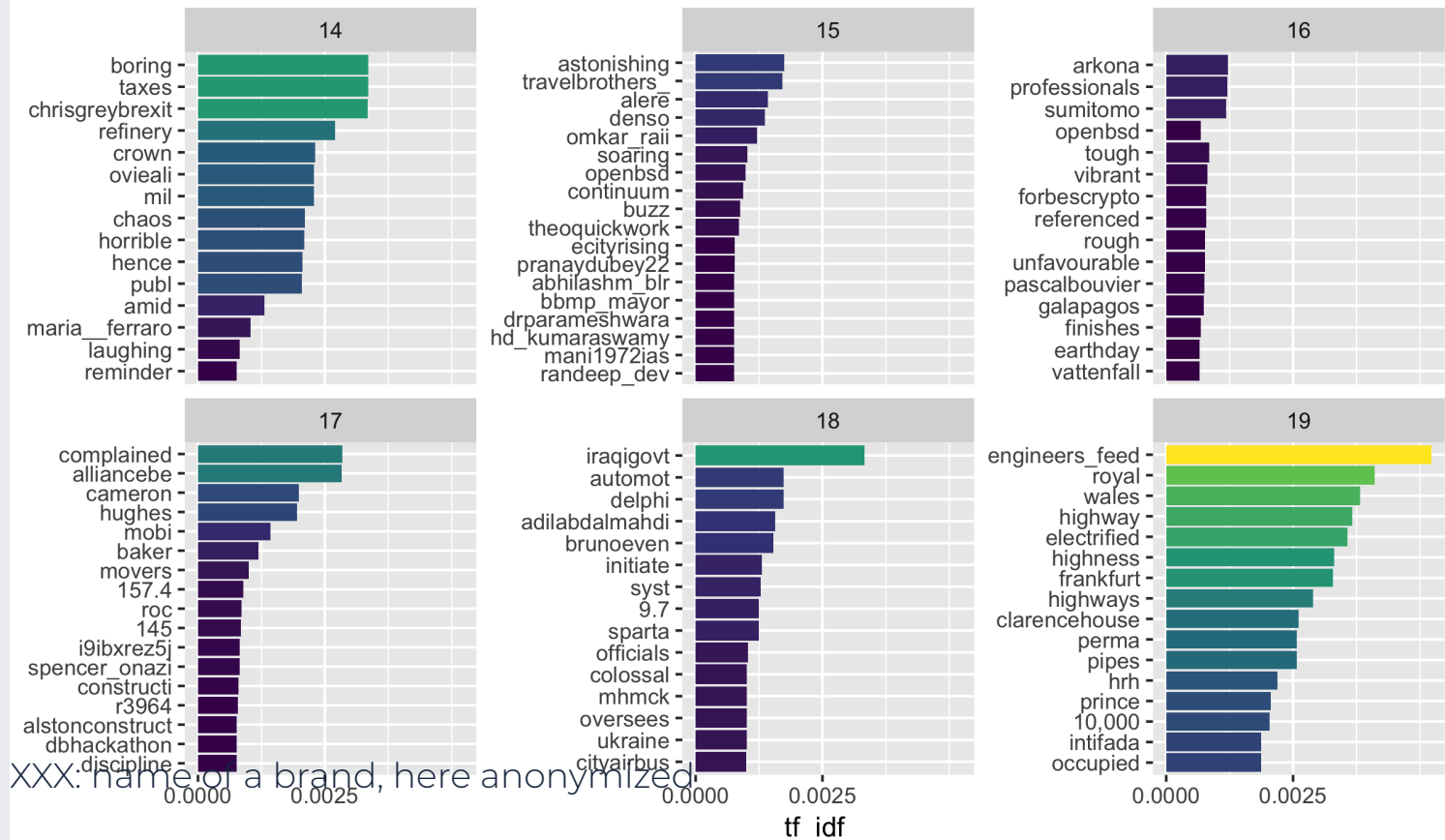
Emotions in tweets with keyword 'XXX'



XXX: name of a brand, here anonymized

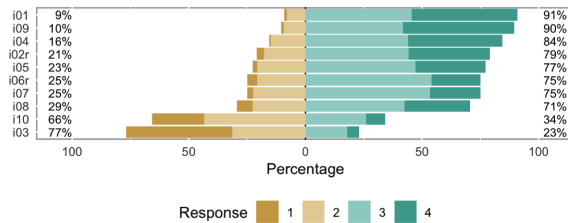
Word frequencies in tweets containing 'XXX'

Word frequencies in Tweets containing 'XXX' (per week)



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

 sauer_sebastian

CC-BY

Reproducibility

- Versions of employed software as of 2019-05-10, running this OS: macOS Mojave 10.14.4.
- Built with [R](#), R version 3.5.3 (2019-03-11), [RStudio](#) 1.2.1335, [xaringan](#), on the shoulders of giants
- Source Code: XXX
- 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, http_1.4.0, icon_0.1.0, ipred_0.9-9, iterators_1.0.10, jsonlite_1.6, knitr_1.22, lattice_0.20-38, lava_1.6.5, lazyeval_0.2.2, lubridate_1.7.4, magrittr_1.5, MASS_7.3-51.1, Matrix_1.2-15, ModelMetrics_1.2.2, modelr_0.1.4, munsell_0.5.0, nlme_3.1-137, 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_4.1-13, 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_2.43-3, tibble_2.1.1, tidyr_0.8.3, tidyselect_0.2.5, tidyverse_1.2.1, timeDate_3043.102, withr_2.1.2, xaringan_0.9, xaringanthemer_0.2.0, xfun_0.6, xml2_1.2.0, yaml_2.2.0
- Last update 2019-05-10