

---

# Medical Drug Review Analysis

---

**Simon Reif**

Matrikelnummer: 5617379

simon.reif@student.uni-tuebingen.de

**Dejana Mandic**

Matrikelnummer: 6024572

dejana.mandic@student.uni-tuebingen.de

## Abstract

In this project, we will be examining the UCI ML Drug Review Dataset (Li, 2019), which contains a list of medical drugs along with the medical condition they are used for, user reviews, grades(rating), and the number of people who found a certain review useful. We want to check how some words (and/or text length) in user reviews are correlated with the overall user grade, and whether that depends on the medical condition the medication is used for. Further, we want to investigate what are the most common words in the reviews people find useful and order the medical conditions and medications by how much people are relying on the reviews. Some other ideas we might try are: predicting the rating using the reviews; sentiment analysis of the reviews (positive/negative) and whether it might be related to the date of the review.

## References

Li, J. (2019). *UCI ML Drug Review dataset - Version 2*. Available at <https://www.kaggle.com/jessicali9530/kuc-hackathon-winter-2018>, version 1.6.0.