

# Data Science – Gruppe Midori

Buchegger / Cardini / Schmocker

# ArffGenerator

- ▶ Lädt die Kommentare (txt)
  - ▶ Stemming, um Worte auf ihren Stamm zurückzuführen
  - ▶ Negationen werden berücksichtigt
- ▶ extrahiert die Features
  - ▶ General Inquirer Lexicon, um Konnotationen der Wörter zu bestimmen
  - ▶ Features = Anzahl Konnotationen (positiv, negativ, erregt, schmerzhaft, ...), welche im Kommentar vorkommen
  - ▶ Gewichtung anhand der «Wortstärke»
- ▶ speichert die Instanzen und Features als ARFF-File ab

# App

- ▶ Baseline Evaluation
- ▶ Generieren eines Naive Bayes Classifiers
- ▶ Tenfold cross-validation (stratified)

# Fazit

- ▶ Eigener ARFF-Generator -> Keine Gute Idee
  - ▶ Schwierig weitere Weka-Filter einfach einzubauen
- ▶ Snowball-Stemmer macht keinen Unterschied

The background features abstract, overlapping geometric shapes in various shades of green, primarily concentrated on the right side of the image. The word "Code" is centered in a large, green, sans-serif font.

# Code

# Fragen