



Sentiment Analysis Project

Simon Schaad, Jan Scheidegger, Nicolas Schmid, 15.1.2018

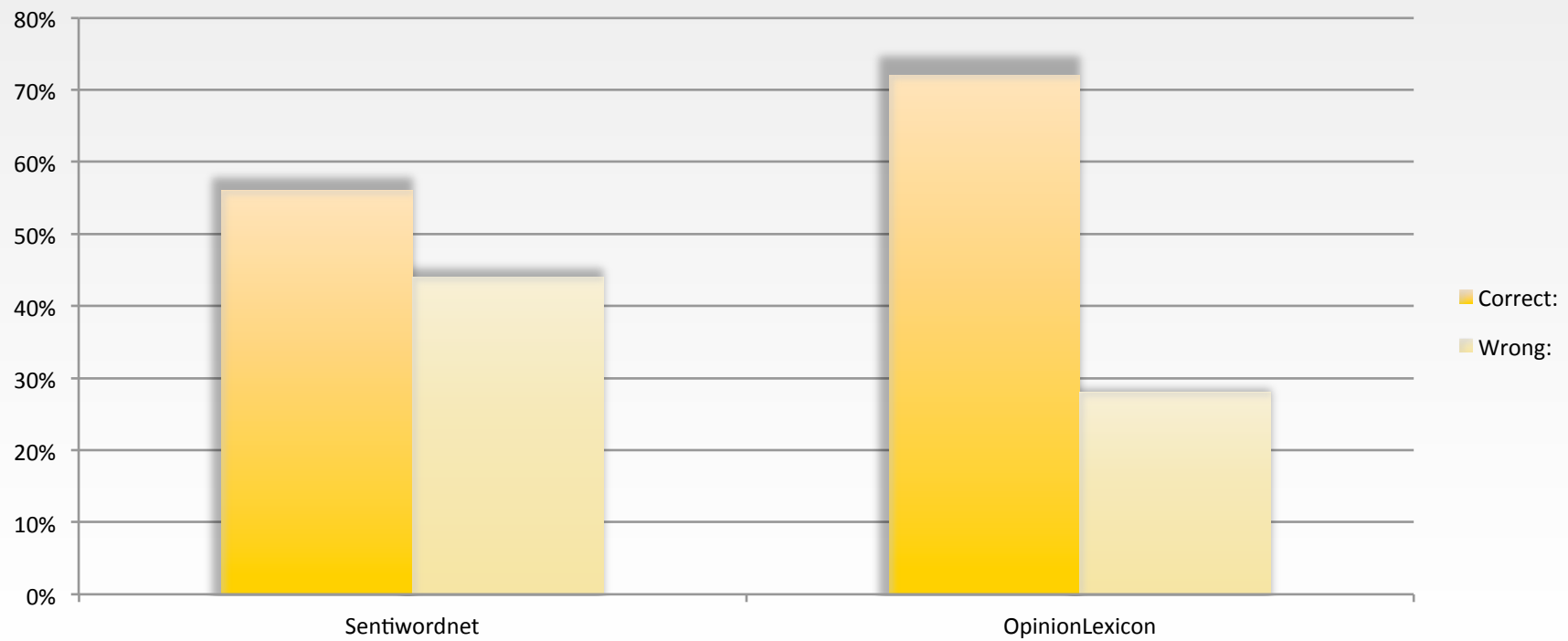


Ansätze / Features

- Baselines mit SentiWordNet & OpinionLexicon
- Stopwordlist
- Div. Tokenizers
- Special Character Replacement
- Negation Filter
- Shortwords Filter
- Unused Char Filter
- Rating Filter
- PorterStemmer
- NGramRainbow
- Weka

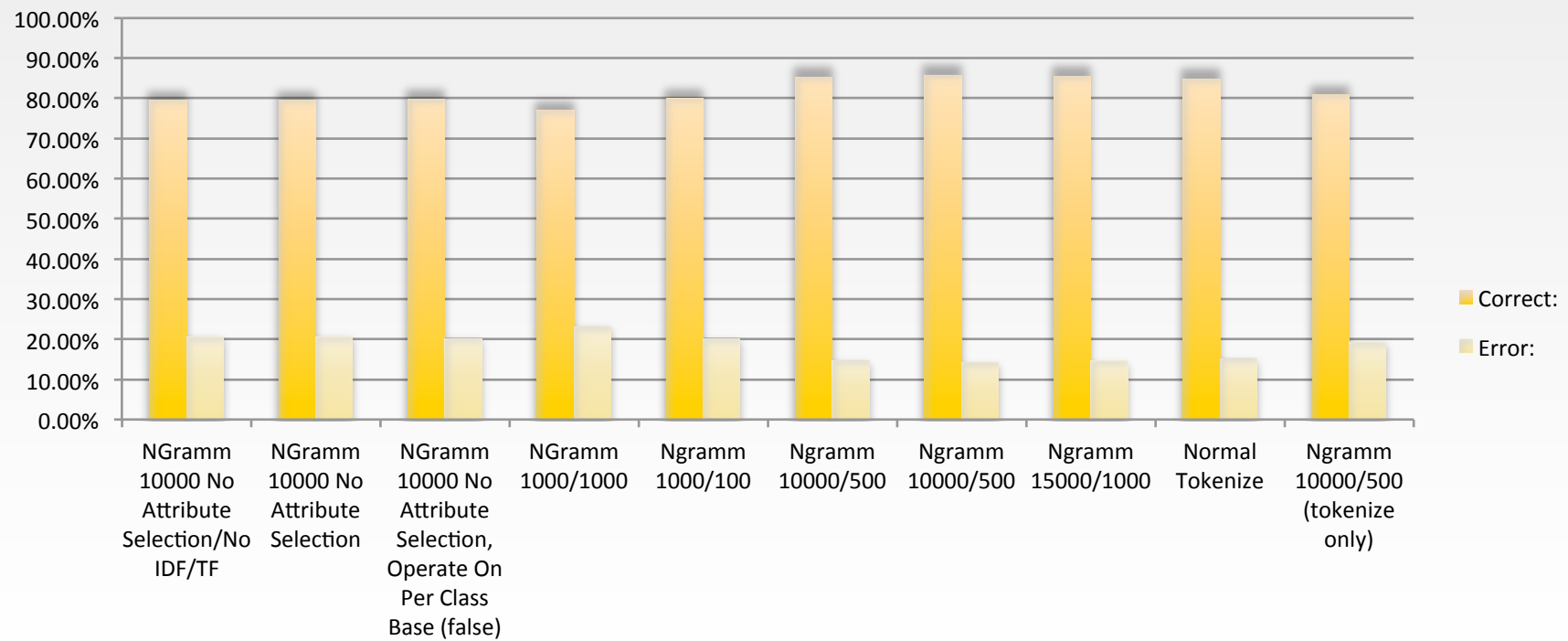


Resultate





Resultate





Lessons Learnt

- Quite tricky
- (Correct) Tokenization very important
- Hard to find features which should outperform „industry standard“
- Lot of testing needed to have solid comparable results