

Predicting swiss reviews helpfulness

Abstract: We classify reviews for Swiss products based on helpful / 'not helpful' flags, while evaluating several models.

Motivation: More than half of the reviews on swiss products do not have helpful scores and end up being ignored or never read.

Classification models:

- Logistic Regression
- Random Forests
- AdaBoost
- SVM (Linear SVC)

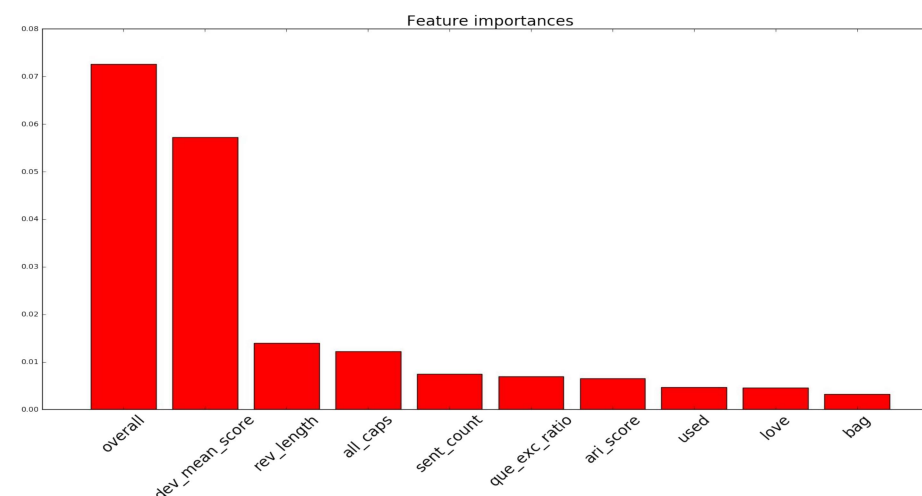
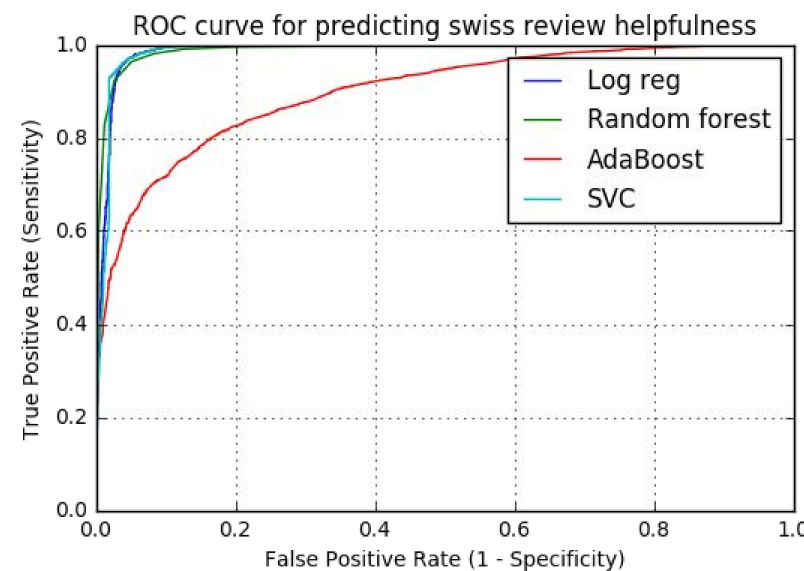
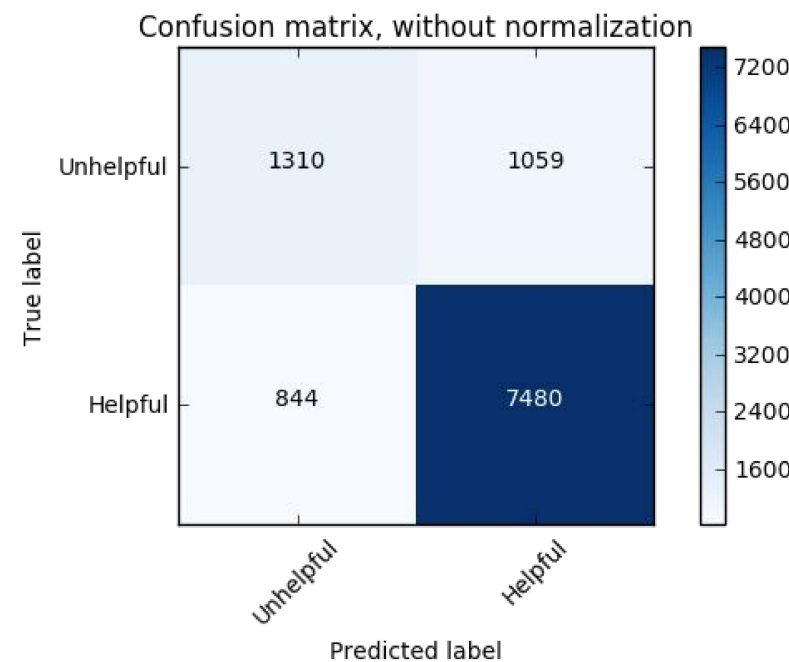
Metrics used:

- Confusion matrix
- ROC curve

Features used:

- Tf-idf
- Rating
- Review length
- Caps ratio
- Sentence count
- '?' and '!' ratio
- ARI readability score
- Deviation from mean score

Number flags > 10



Number flags > 5

