

# Authorship Attribution

*by Jamal Abdul Nasir,  
Nico Gornitz, Ulf  
Brefeld*

Наташа Акутина,  
Ира Павлова

# Что использовали авторы

- по 10 текстов 10-ти авторов
- 90% на обучение и 10% на тест
- SVM
  - POS
  - 3-letter suffixes
  - bag-of-words (tf-idf)
  - частотность служебных слов (например: авторский стиль по частоте оборотов типа “bag of words” vs. “words bag”)



- мы выкачали все рецензии на музыкальные альбомы с сайта Rolling Stone (<http://www.rollingstone.ru/music/review/>)
- мы отобрали 586 текстов 9-ти авторов

# Для нашего обучения мы будем использовать:

метод опорных векторов, с учетом следующих признаков:

- на первом этапе
  - POS
  - bag-of-words
- на следующем этапе
  - служебные слова