

Author Profiling

Borna Sirovica, Filip Zelić, Ivan-Dominik Ljubičić

Fakultet elektrotehnike i računarstva

June 13, 2016

Content

- 1 Project topic
- 2 Dataset
- 3 Preprocessing
- 4 Feature extraction
- 5 Supervised learning model
- 6 Results

Author Profiling

- Identifying information about an author by analyzing their writing style
- Distinguishes between classes of authors studying their sociolect aspect
- Growing importance in forensics, security, marketing, and social networks



Our task

- Predicting author profiling aspects
 - Demographics (Classification)
 - age
 - gender
 - Big Five personality traits (Regression)
 - extroversion
 - stability
 - agreeableness
 - conscientiousness
 - openness

- English tweets dataset from PAN 2015 competition
 - XML format, 152 users
 - Age and gender labels
 - 18-24, 25-34 34-49, 50+
 - M, F
 - Personality traits
 - normalized numeric rating in $[-0.5, 0.5]$ range

Preprocessing

- Creating a document of tweets for each user
- Cleaning tweets from
 - XML tags
 - hashtags
 - @mentions
 - URLs
 - duplicates

Features

- Ovdje opisati što o featurima

Feature extraction

Models

Odabri modela.

Rezultati.

Implementation

- Python Python
- Sklearn Sklearn

Conclusion

- Zaključak rada.

Thank you for attention.

Questions, comments ?