

# Authorship analysis

Roberto Deresu

Facultatea de Matematica si Informatica



# Dataset

- The Federalist Papers Dataset
- PAN12 - Text Authorship Attribution Dataset

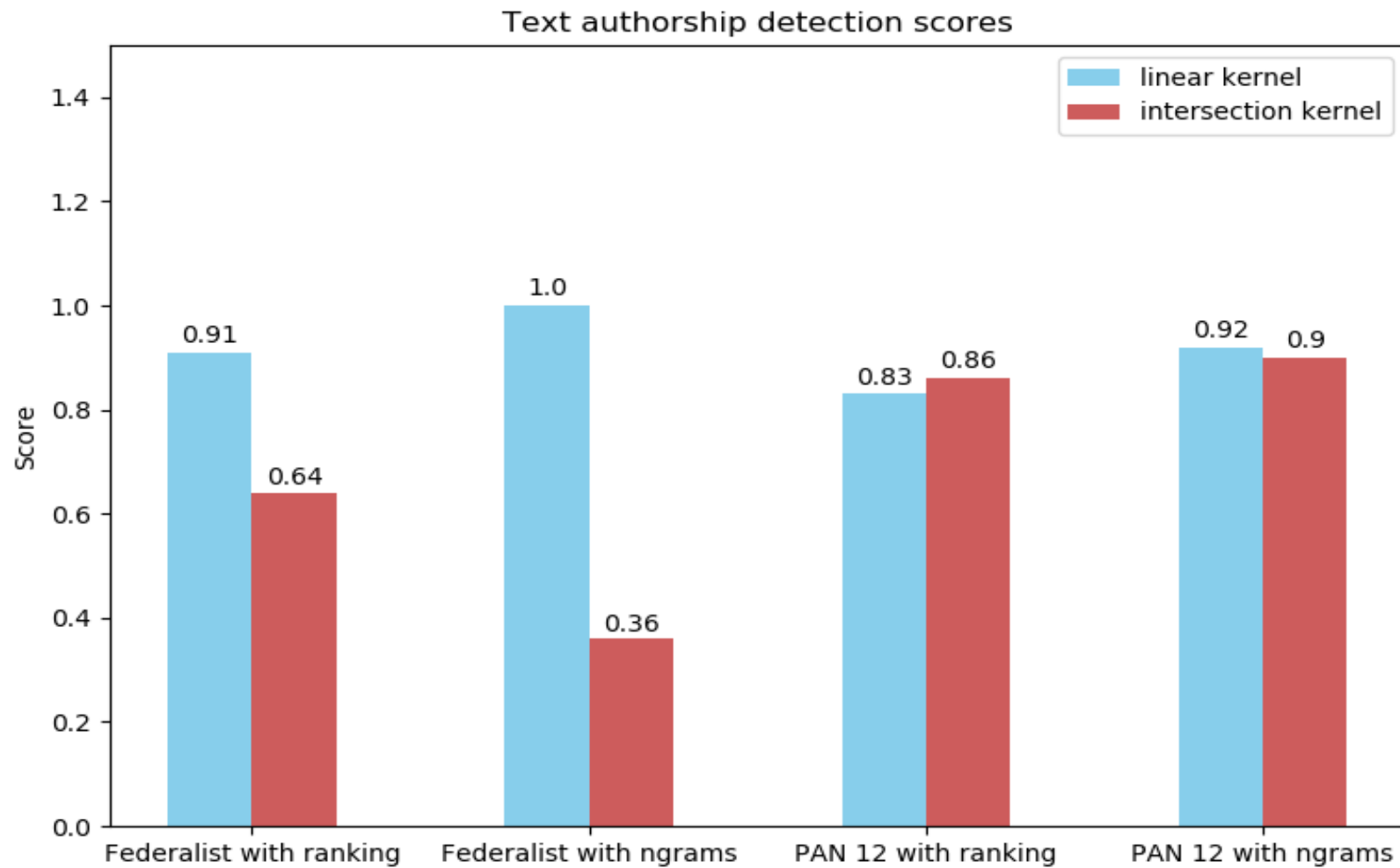
# Features

- Ranking distance representation of each document, with function words: Mosteller and Wallace or nltk english
- N-grams representation of each document. N-grams range: 3-4, frequency threshold: 100

# String-kernels

- Linear kernel
- Intersection kernel
- Binary kernel

# Results



THANK YOU!