THE ANDREW & ERNA VITERBI
FACULTY OF
ELECTRICAL
ENGINEERING

SIPL
Signal and Image Processing Lab

TECHNION
Israel Institute
of Technology

Identification of an Anonymous Reviewer of an Academic Paper

Lior Kiassi, Roy Hachnochi, Supervised by Pavel Lifshits

Introduction

- Commonly, as part of the process of publishing a scientific paper, the work undergoes peer review.
- “Double Blind” – the author and the reviewer are anonymous to each other.
- Today, due to transparency reasons, several well-known journals publish their reviews.

Goals

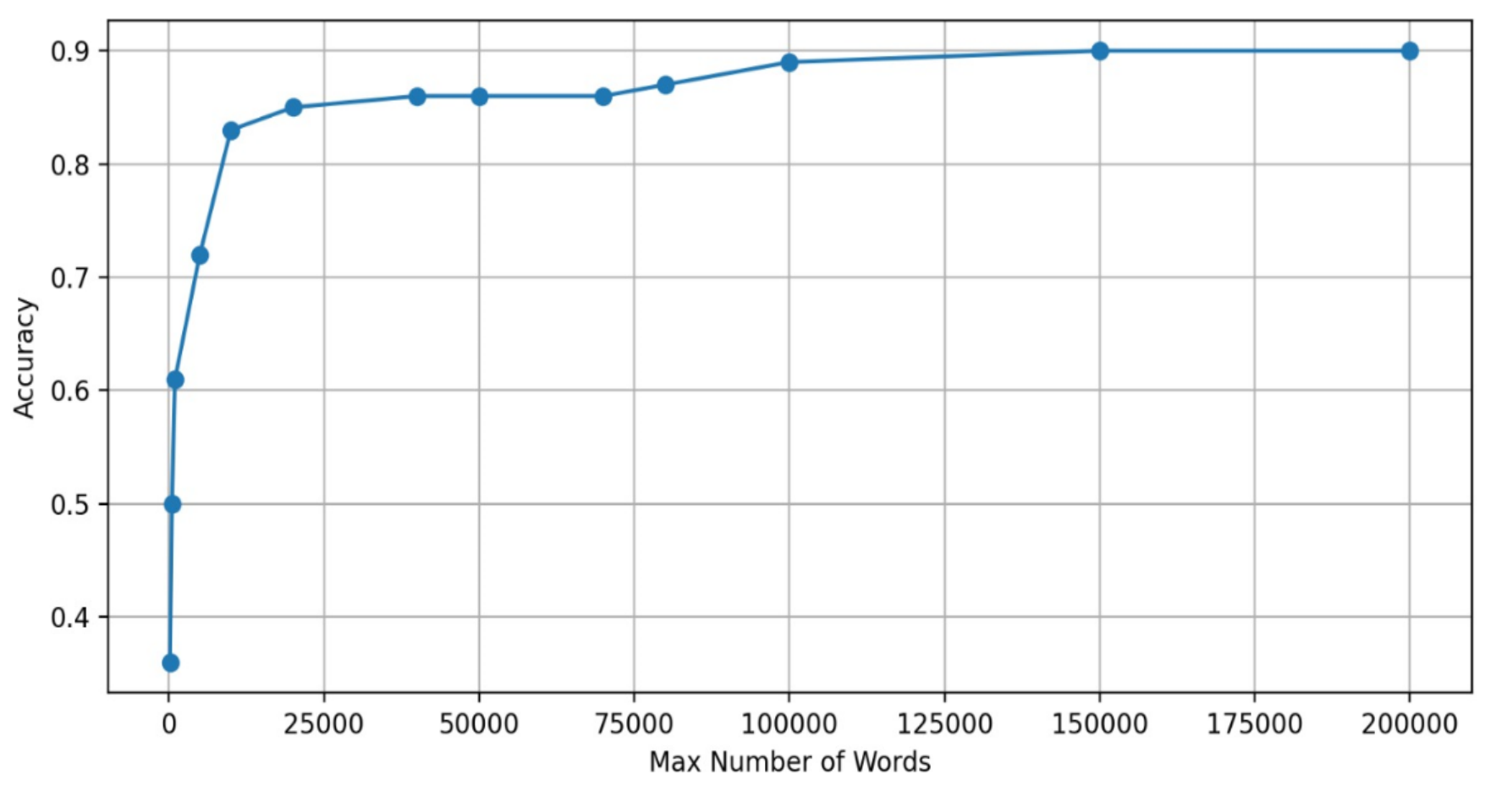
- Study the domain of a new problem – reviews on academic articles.
- Research how to perform authorship attribution on this domain by learning a different domain.
- Stepping-stone towards the following:
    - Obfuscation of reviews’ authors.
    - Dealing with the cross-domain problem – learning on articles and inferring on reviews.

Challenges

- No existing labeled ground truth.
- Reviews are short texts.
- Reviews and academic papers are topic related.
- Cross domain work.
- Most articles are written by multiple authors.

Toy Problem

- Small dataset: 100 texts – 10 Authors x 10 books (Guttenberg project).
- **Goal: create an author classifier (90% Acc.).**
- Accuracy vs. text length:

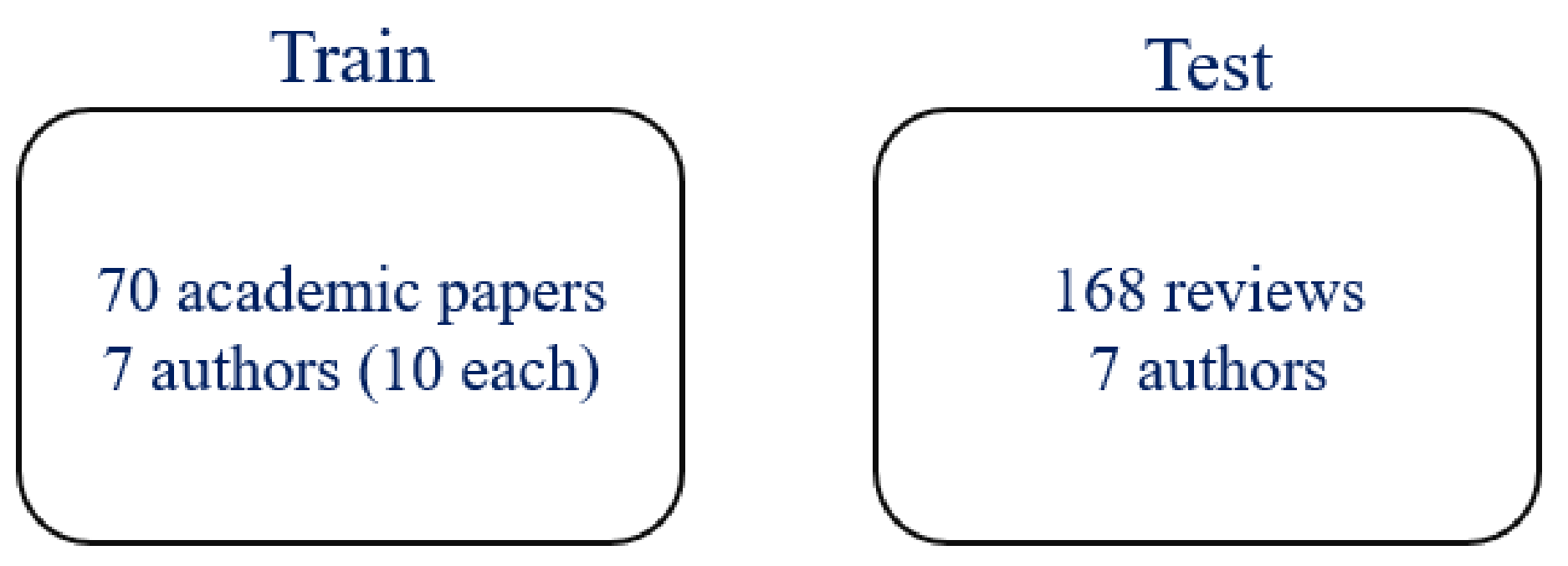


Dataset

Challenges:

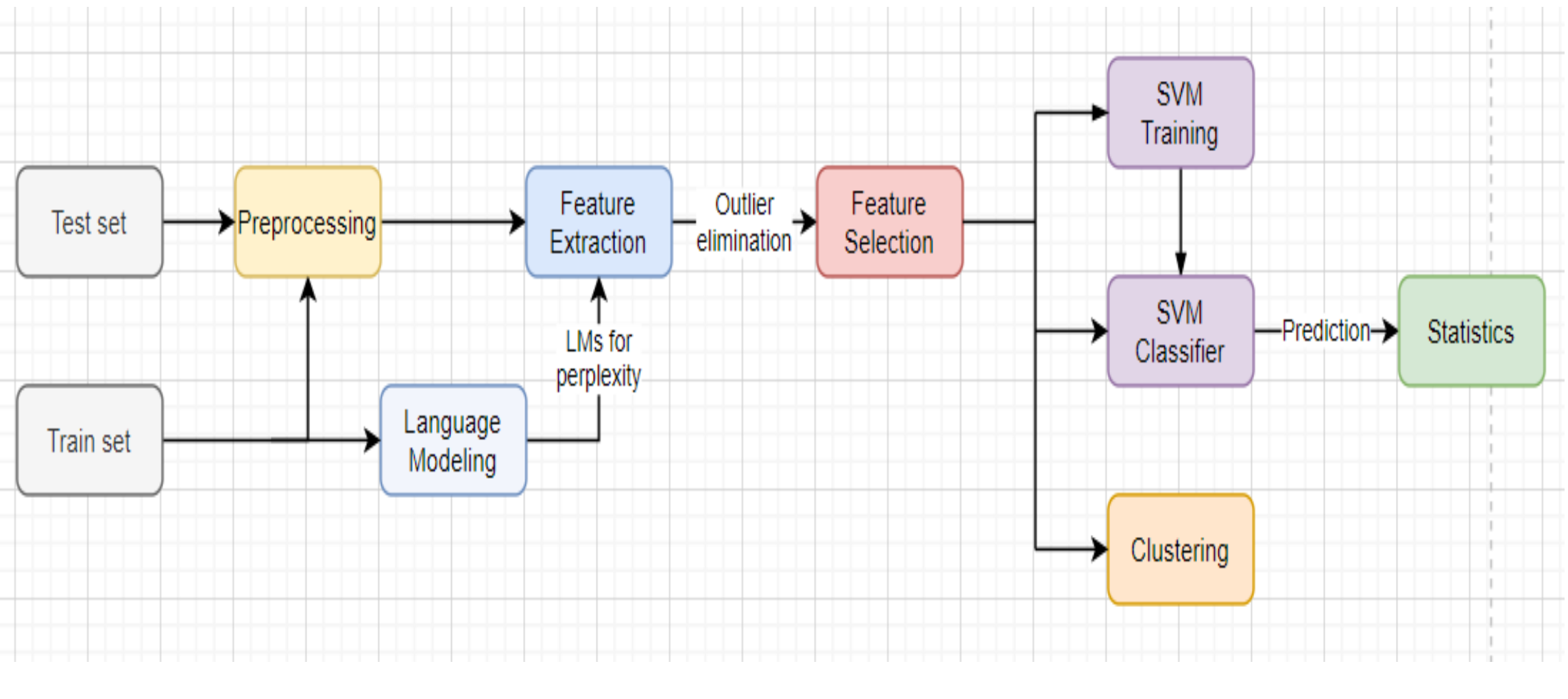
- Hard to find (especially labeled ones).
- Review’s authors write only few reviews in total.
- Short texts without meaningful context.
- Different text structures – hard for automation.
- Difficulty to obtain papers of a specific person.

Structure:


Train
70 academic papers
7 authors (10 each)
Test
168 reviews
7 authors

- Reviews - BMJ (British Medical Journal).
- Articles - various sources.

Project Diagram



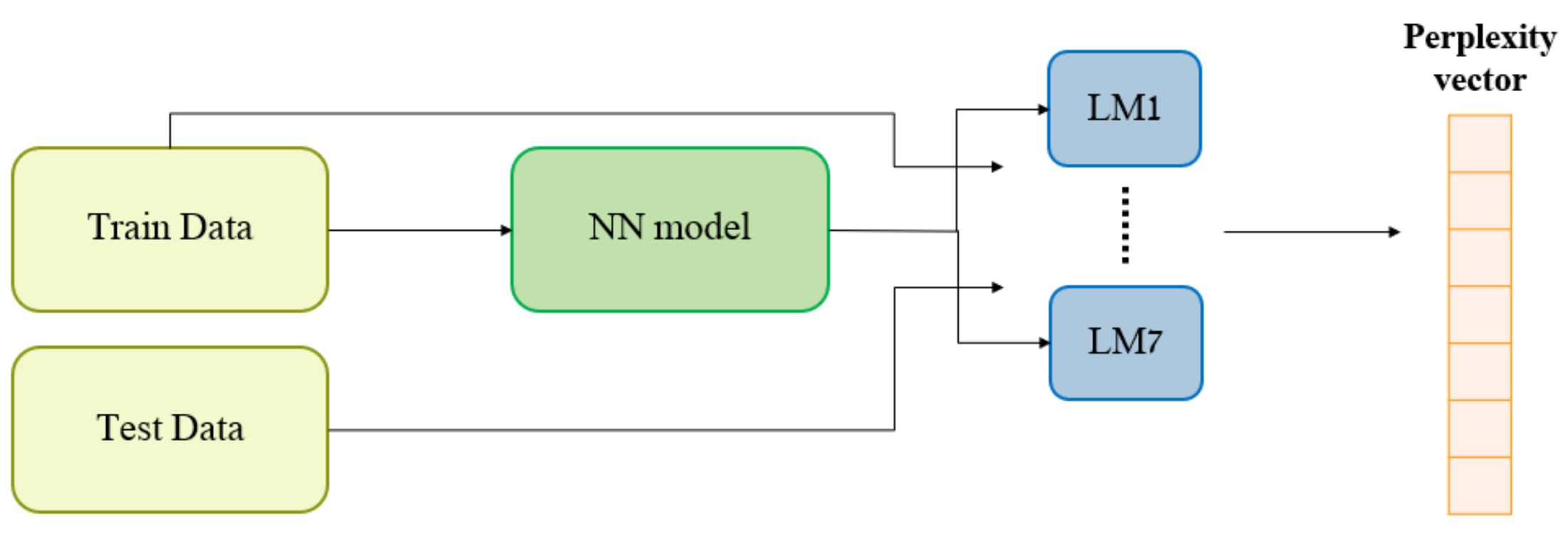
Preprocessing

- Tokenizing.
- Ignoring stop words.
- Lemmatization.
- Replacing numbers and signs with ‘<UNK>’.

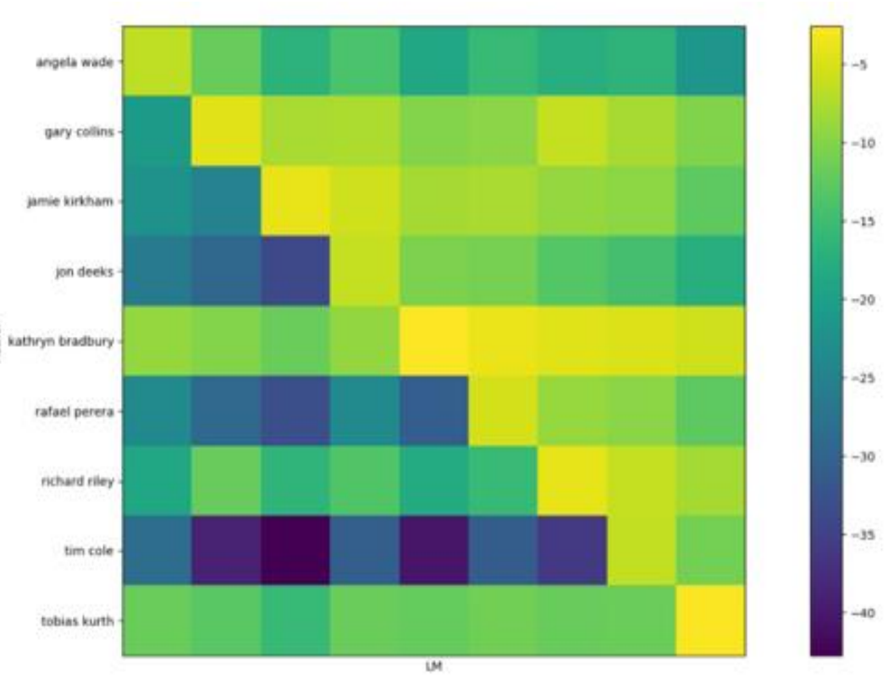
Features

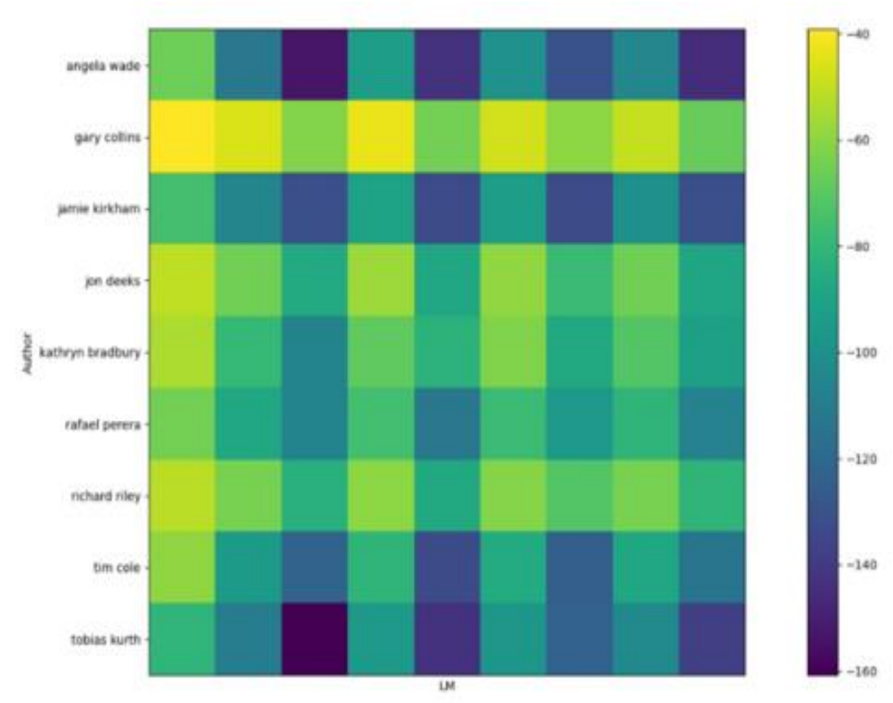
- n-gram histograms:
    - $n \in \{1,2,3,4,5\}$
    - TF-IDF normalization: $h[i] = \frac{n_i}{N} \cdot \log\left(\frac{D}{N_i}\right)$
- Average word length.
- Number of words.
- Average number of words in a sentence.
- Histogram of the following punctuation signs: , ; : -

Features – Language Models



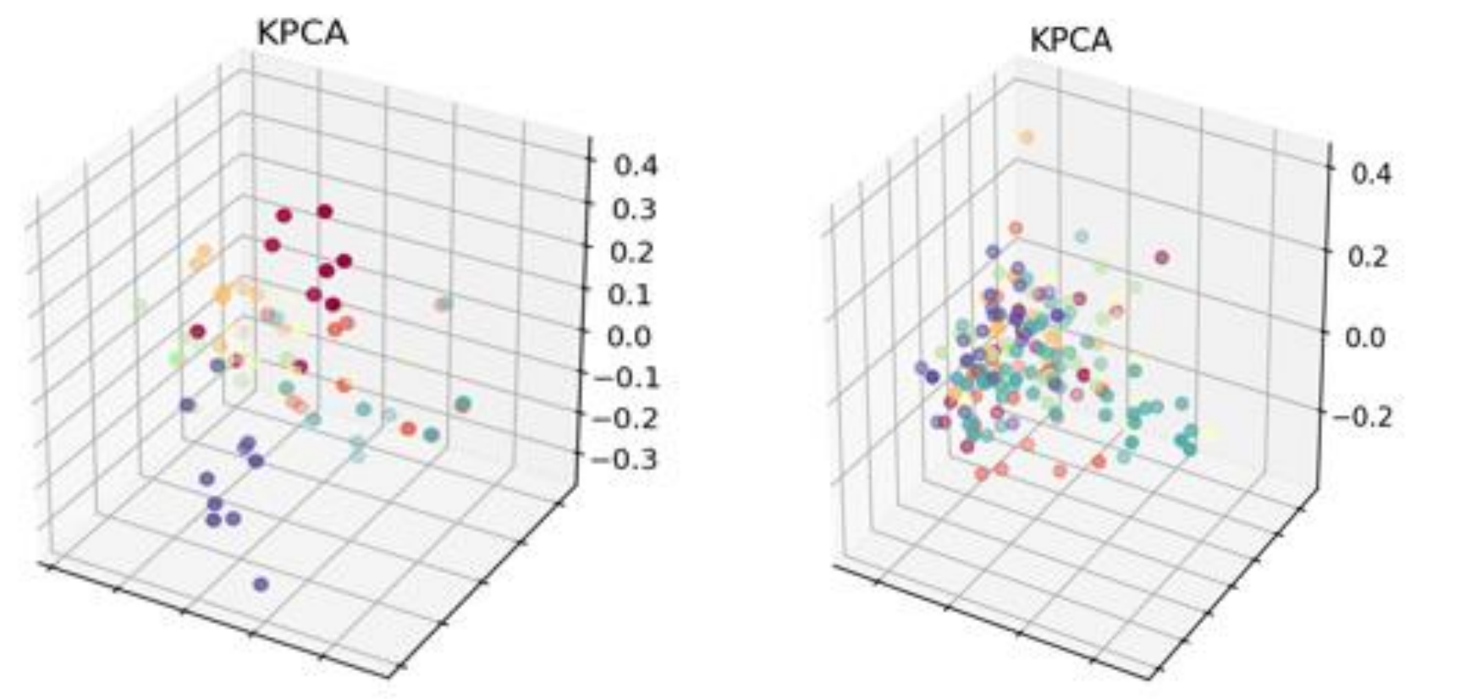
- Language model for each author.
- Calculates the probability for the next word.
- LSTM – Huggingfaces’ Transformers with GPT2.
- Perplexity as feature: $perp(x) = e^{-\frac{1}{N}\sum_{i=1}^{N}\log(P_\theta(x_i))}$
- Perplxity indicates the likelihood of a text to be associated with a language model.
- Gives understanding of vocabulary, writing style and more.


Articles - Articles


Articles - Reviews

Clustering

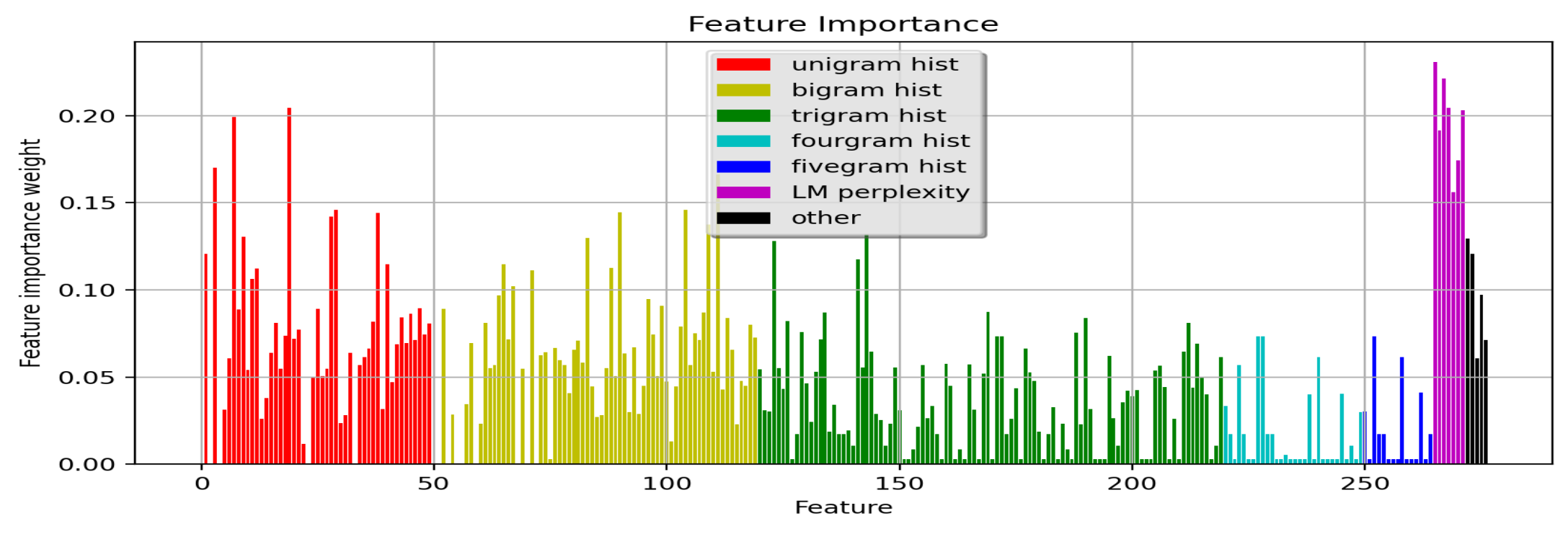
- Finding outliers in dataset – OPTICS algorithm.
- Samples visualization.
- Manifold learning – learning the feature space of the train set and the test set.



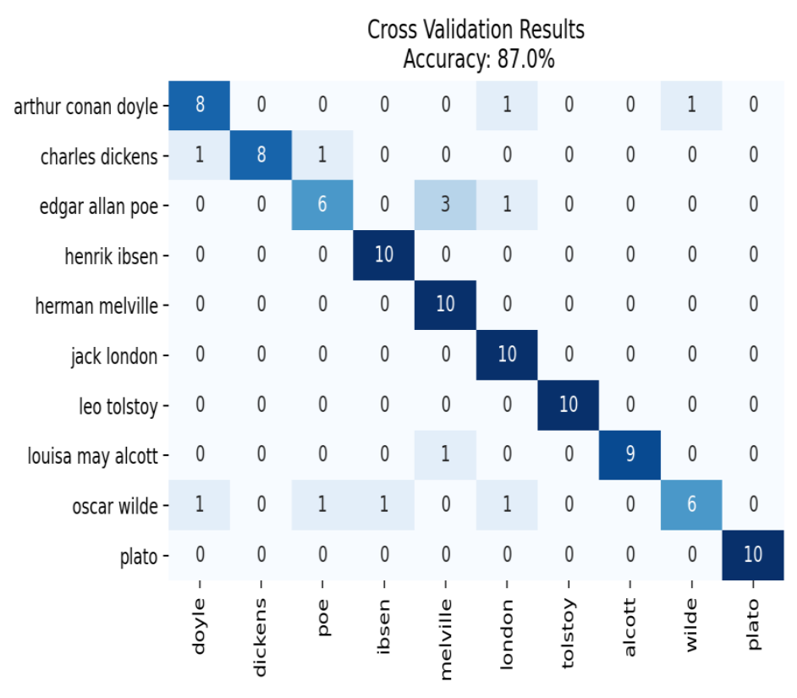
Articles
Reviews

Feature Selection

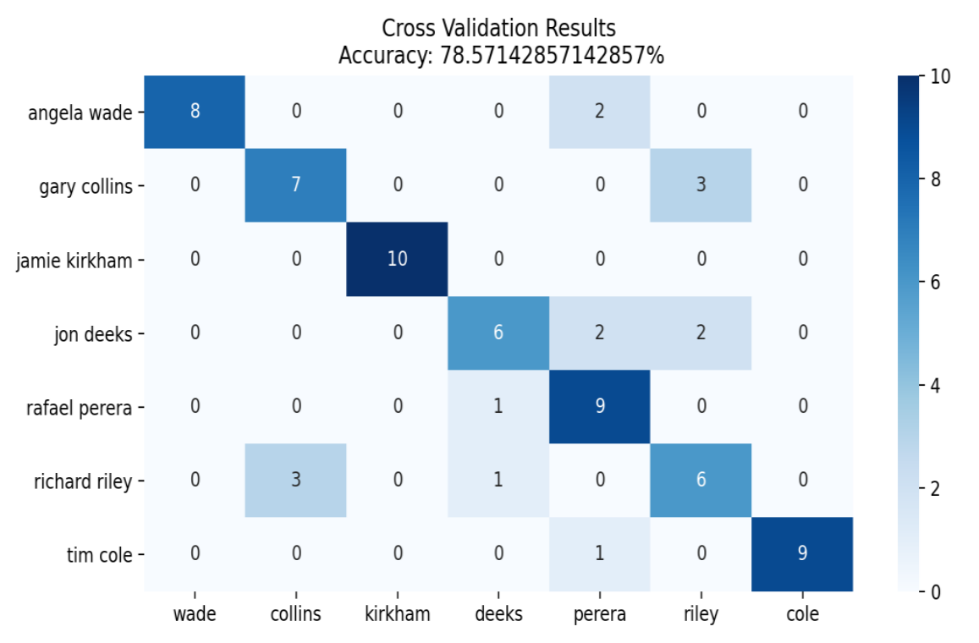
- Eliminate bad features (for both domains) with reliefF algorithm.
- For example – articles:



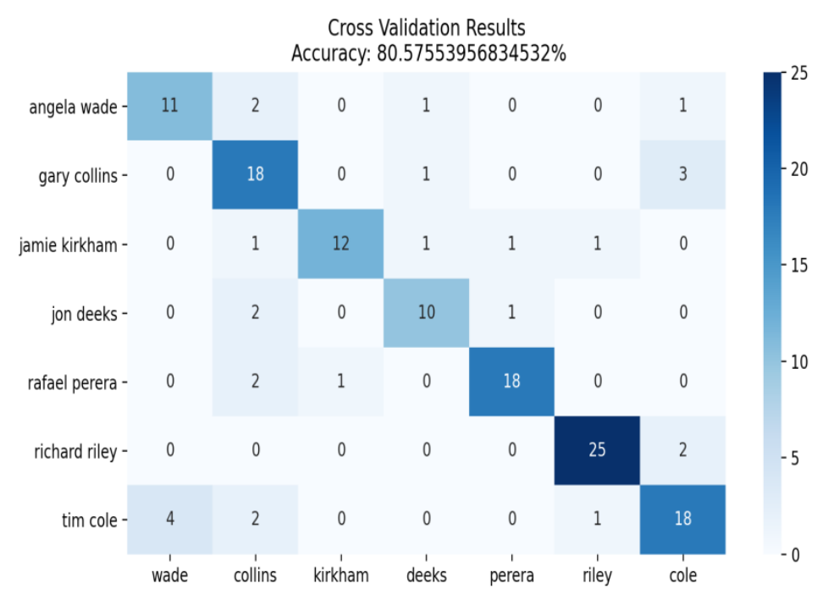
Results



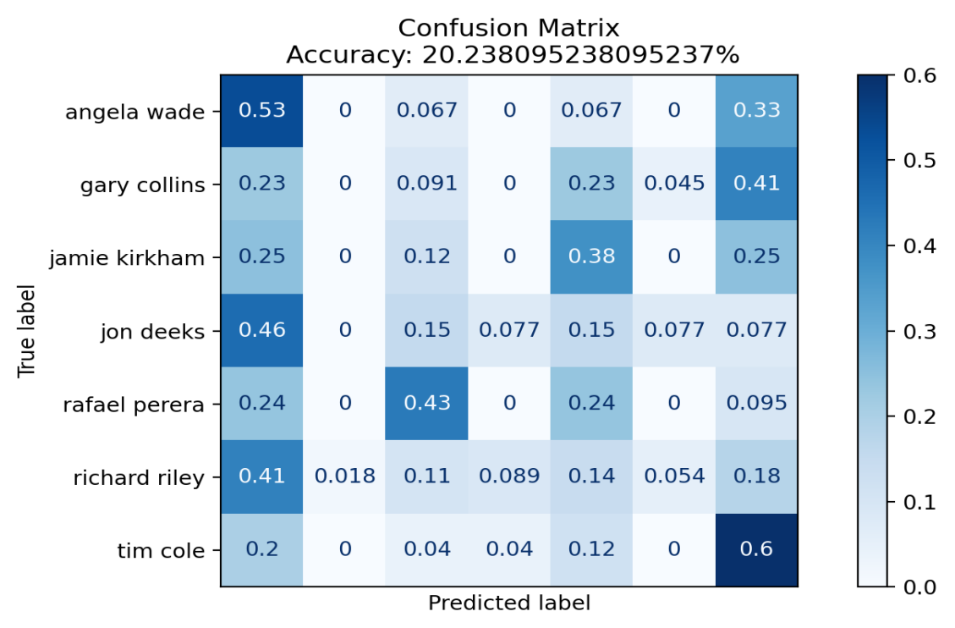
Toy problem



Articles - Articles



Reviews - Reviews



Articles - Reviews

- Indicates on the stability of our model for various problems.
- Stepping-stone for possible future work.
- Cross-domain problem hasn’t been dealt with – for future work.

Conclusions

- Good results on a naturally challenging domain.
- Future work – domain adaptation.
- The quality and size of the dataset have impact on the results.
- Text length may have a great impact on results.
- Using language models and perplexity is a highly valuable feature for this task.
- Feature selection is important when dealing with high feature dimension.

November 2020