Poem Classification
Machine Learning for Natural Language Processing 2020

Emeline Maire
emeline.maire@ensae.fr

Paul Petit
paul.petit@ensae.fr

Abstract

Using the poetry foundation database, the aim of our project will be to classify poems between five different authors. We will be using different techniques to find out which classification algorithm works better. We will use a classification algorithm based on the vocabulary with an TF-IDF algorithm and then also a classification baed on the metrics of the poems. [1]

1 Problem Framing

Using the poetry foundation data set, our goal will be to classify poems given their author. In order to have a rather balanced data set we only keep authors with at least 50 poems in the database.

Table 1: Number of poems by author

Name	Number
Willam Shakespear	85
Lord Alfred Tennyson	78
Rae Armantrout	62
William Wordsworth	59
Emily Dickinson	57
Total	341

In order to classify these poems by authors, we will compare different techniques of classification.

2 Experiments Protocol

To lead our analysis we will use three different techniques. First, we will based our classification algorithm on the vocabulary used by the authors, with TF-IDF algorithm. Then, we will use a comparison of the authors based on the rhythm of the poems.

[1] https://github.com/Emeline-Maire/NLP-project.git

3 Results

3.1 TF-IDF

The first model used in this analysis is a random forest classifier based on the vocabulary used by each author. To do so, we used the TF-IDF vectorization of the poems of our dataset. The TF-IDF method uses the frequency of a token in a document to measure its importance ; in this case we look at the importance of some words in poems, and more generally see if there are large differences in vocabulary between authors that would allow us to predict the authors based only on the words used and thair frequency.

	precision	recall	f1-score	support
Alfred, Lord Tennyson	0.89	0.67	0.76	24
Emily Dickinson	0.52	0.72	0.60	18
Rae Armantrout	0.87	0.68	0.76	19
William Shakespeare	0.55	0.85	0.67	26
William Wordsworth	1.00	0.39	0.56	18
accuracy			0.68	105
macro avg	0.77	0.66	0.67	105
weighted avg	0.76	0.68	0.68	105

Figure 1: Results of the TF-IDF prediction

The results are already quite satisfying as our average f1-score is of 0.68. However, our model tends to over predict William Shakespeare and struggles to predict Wordsworth correctly. Hence, we try with another model using word embedding and neural networks.

3.2 Word embedding & neural network

For this model, we used pre-trained embeddings and used the TreebankWordTokenizer. Once the data preparation is done, we use a deep learning model to predict the author of the poems. Our neural network is made of one hidden layer.

This model is less accurate than the TF-IDF one as our average f1-score is of 0.56 (see table in the appendix). The model also tends to over predict

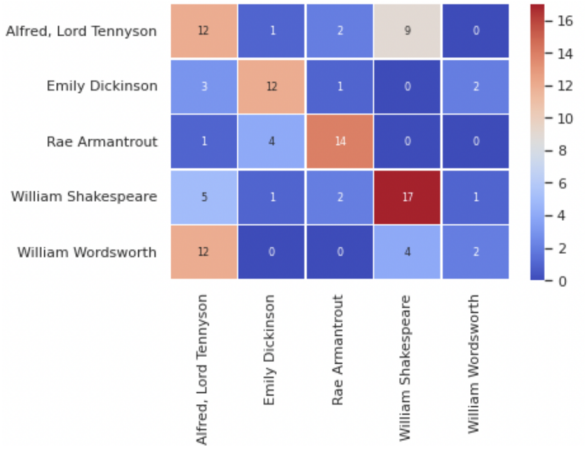

Figure 2: Confusion Matrix for the neural network

Shakespeare and struggles to predict Wordsworth. In addition, the model struggles to classify correctly the poems of Lord Tennyson. As a result we can imagine that Lord Tennyson had a very peculiar vocabulary as the TF-IDF model managed to classify his poems better.

3.3 BERT

The third model used in our analysis, is the pre-trained one : BERT - more precisely ber-tiny which is faster to train.

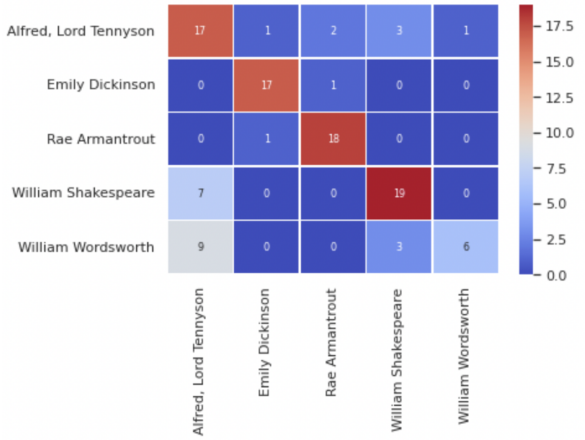

Figure 3: Confusion Matrix BERT

This model is better than the neural network one and also slightly more efficient than the TF-IDF one. This model struggles to classify Wordsworth's poems. Indeed, the average f-score is of 0.73, and the f1-score for Wordsworth is of 0.48 which is worse than the TF-IDF model for this author.

3.4 Rhythm analysis

Lastly, we implemented a model which takes into account the rhymes of the poem and the rhythm, with the number of strophes and lines. For the rhyme we only look at the first three lines of the poem.

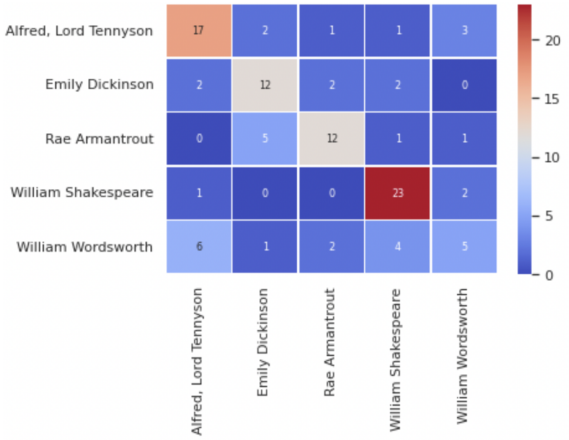

Figure 4: Confusion Matrix for the rhyme and rhythm model

This very simple model still manages to classify rather well the poems, still Wordsworth appears very difficult to predict. The average f1-score is very close to the TF-IDF one.

4 Discussion/Conclusion

We used four different models to test which one would manage to classify best the poems. In the end, we have that the best model for our data set is the TF-IDF. However, across all models, the poems of Wordsworth were quite poorly classified. The relatively low scores obtained for his poems across the models might be due to the fact that he had less poems in the data set than others, plus his writings might not be as singular as some others. Indeed, Rae Armantrout might not have so much more poems in the dataset but, as she is a contemporary writer (born in 1947), she has quite a unique style, especially compared to the four other authors in the analysis. Also Emily Dickinson is quite unique as she is one of the first American poetess. Among the first generations of American writer, there was a very strong will to cut off from the British ones, with a will to invent a new style, sometimes new words. Another conclusion which can be drawn from this analysis is that classic models are not very efficient to classify poems as it is a specific form of literature and models that take these specificity into account are better.

References

Mans Agirrezabal, Manex & Hulden. Machine learning for metrical analysis of english poetry.

Stefan Niculescu, Irina-Diana & Trausan-Matu. Rhythm analysis of texts using natural language processing.

Marmik Pandya. 2016. Nlp based poetry analysis and generation. *Northeastern University*.

Appendix

	precision	recall	f1-score	support
Alfred, Lord Tennyson	0.36	0.50	0.42	24
Emily Dickinson	0.67	0.67	0.67	18
Rae Armantrout	0.74	0.74	0.74	19
William Shakespeare	0.57	0.65	0.61	26
William Wordsworth	0.40	0.11	0.17	18
accuracy			0.54	105
macro avg	0.55	0.53	0.52	105
weighted avg	0.54	0.54	0.52	105

Figure 5: Scores of the neural network model

	precision	recall	f1-score	support
Alfred, Lord Tennyson	0.52	0.71	0.60	24
Emily Dickinson	0.89	0.94	0.92	18
Rae Armantrout	0.86	0.95	0.90	19
William Shakespeare	0.76	0.73	0.75	26
William Wordsworth	0.86	0.33	0.48	18
accuracy			0.73	105
macro avg	0.78	0.73	0.73	105
weighted avg	0.76	0.73	0.72	105

Figure 6: Scores of the BERT model

	precision	recall	f1-score	support
Alfred, Lord Tennyson	0.65	0.71	0.68	24
Emily Dickinson	0.60	0.67	0.63	18
Rae Armantrout	0.71	0.63	0.67	19
William Shakespeare	0.74	0.88	0.81	26
William Wordsworth	0.45	0.28	0.34	18
accuracy			0.66	105
macro avg	0.63	0.63	0.63	105
weighted avg	0.64	0.66	0.64	105

Figure 7: Scores for the rhyme and rhythm model