SemEval 2018 Task #7
"Semantic Relation Extraction and Classification in Scientific Papers"

1. Baseline.

In order to be able to compare proposed solution to other system, some reference solution needed to be selected. As a baseline, classifier model which predicts every incoming instance as a majority class (relation type "*usage*") was selected. Results from the selected baseline are covered in Section 3, while results obtained from the proposed solution are covered in Section 4.

2. Evaluation metrics.

For the SemEval Task 7, the following evaluation metrics were selected:

- Accuracy, which corresponds to the ratio of correctly classified instances to the total number of instances that were supposed to be correctly classified as seen in the equation below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- F-Measure with macro average in which precision and recall is averaged between classes

$$Fmeasure = 2 * \frac{precision * recall}{precision + recall}$$

Additionally, k-Fold Cross Validation was performed in order to achieve reliable results of performance obtained from the proposed solution.

3. Baseline results.

Results for a baseline solution for the old dataset are covered in the *Table 1*.

Table 1 Results for the baseline on the old dataset.

	Development set [%]	**Test set [%]**
Accuracy	37.84	40.0
F-measure (macro avg)	10.98	11.43

Results for a baseline solution for the new dataset are covered in the *Table 2*.

Table 2 Results for the baseline on the new dataset.

	Split 60/40 [%]	**kCV (k = 5) [%]**
Accuracy	39.51	39.35
F-measure (macro avg)	9.44	9.41

4. Results obtained from the proposed solution.

Following results are obtained from the framework designed with SemEval 2018 Task 7 in mind with using only features from Feature Extractor 1 which are listed below. As a classification algorithm, the Decision Tree from Scikit-Learn library was used to conduct the experiment.

Features describing each instance:

- Word distance between entities
- Part-of-Speech tags of both entities
- Rule based feature that allows to emphasize minority class

Results for the old dataset are presented in *Table 3*.

Table 3 Results for the proposed solution on the old dataset.

	Development set [%]	**Test set [%]**
Accuracy	70.27	52.0
F-measure (macro avg)	68.75	47.02

Results for the new dataset are presented in *Table 4*.

Table 4 Results for the proposed solution on the new dataset.

	Split 60/40 [%]	**kCV (k = 5) [%]**
Accuracy	53.77	56.13
F-measure (macro avg)	40.67	44.39