

Letter dataset: Results of Text mining project 2020

Your name and UCO

PA164 and KD Lab FI MU Brno

Abstract. This work addresses the problem of ...

1 Data set and task description

1.1 Data set

link to the data set, number of instances, features, missing values etc , data preparation (e.g. transformation into form)

1.2 Data set reduction

Feature reduction

introduce the learning curve, explain what number of feature you will use and why

Other reduction

voluntary; only in the case that you need eg. limit the number of instances

2 Description of the method used

2.1 General

what classifiers (CLF) and what outlier detection (OD) methods you used. Do not describe them.

2.2 Document-term matrix

what representation (pre-processing, PP) you used (very likely binary, TF, TF/IDF)

3 Results

3.1 Overview

general description of your results

3.2 Comparison of different combinations of (PP+OD+CLF)

including graphs and statistical tests.

Which combination was the best in term of accuracy.

Is that combination PP+OD+CLF significantly better that the others? <https://is.muni.cz/auth/el/fi/podzi>

3.3 Discussion

including discussion of the outliers that you found interesting

4 Conclusion

5 References

only if you employed something uncommon.