Supplementary Material for
On the Evaluation of Unsupervised Outlier Detection: Measures, Datasets, and an Empirical Study
by G. O. Campos, A. Zimek, J. Sander, R. J. G. B. Campello, B. Micenková, E. Schubert, I. Assent and M. E. Houle
Data Mining and Knowledge Discovery 30(4): 891-927, 2016, DOI: 10.1007/s10618-015-0444-8

PageBlocks

The data set contains information about different types of blocks in document pages. The task of distinguishing them is an essential step in document analysis, namely to separate text from pictures or graphics. If the block content is text, it was labeled here as inlier, otherwise it was labeled as outlier.

Download all data set variants (14.6 MB). Access original data (page-blocks.data.Z)

- PageBlocks (2% of outliers version#01)
- PageBlocks (2% of outliers version#02)
- PageBlocks (2% of outliers version#03)
- PageBlocks (2% of outliers version#04)
- PageBlocks (2% of outliers version#05)
- PageBlocks (2% of outliers version#06)
- PageBlocks (2% of outliers version#07)
- PageBlocks (2% of outliers version#08)
- PageBlocks (2% of outliers version#09)
- PageBlocks (2% of outliers version#10)
- PageBlocks (5% of outliers version#01)
- PageBlocks (5% of outliers version#02)
- PageBlocks (5% of outliers version#03)
- PageBlocks (5% of outliers version#04)
- PageBlocks (5% of outliers version#05)
- PageBlocks (5% of outliers version#06)
- PageBlocks (5% of outliers version#07)
- PageBlocks (5% of outliers version#08)
- PageBlocks (5% of outliers version#09)
- PageBlocks (5% of outliers version#10)
- PageBlocks (9% of outliers)
- PageBlocks (10% of outliers)

File generated: 2016-07-05T21:48:51