# Data

## Samples

* repo2018\_test – test set of the classifier. A random 5,000 samples from the large active repositories of 2018. Taken from "Which refactoring reduces bug rate?" and detailed there. Named there as labels\commits\_updated2.csv