

Dataset type	Main	Dry-run
Number of categories	12,294	1,139
Number of training instances	93,805	4,463
Number of validation instances	34,905	1,860
Number of test instances	34,880	1,858