

# Large Movie Review Dataset

This is a dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. We provide a set of 25,000 highly polar movie reviews for training, and 25,000 for testing. There is additional unlabeled data for use as well. Raw text and already processed bag of words formats are provided. See the README file contained in the release for more details.

[Large Movie Review Dataset v1.0](#)

When using this dataset, please cite our ACL 2011 paper [\[bib\]](#).

## Contact

For comments or questions on the dataset please contact [Andrew Maas](#). As you publish papers using the dataset please notify us so we can post a link on this page.

## Publications Using the Dataset

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. (2011). [Learning Word Vectors for Sentiment Analysis](#). *The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*.