

Better sentiment classification

Logistic regression over bag of 1,2-grams with TF-IDF

- Accuracy on test set: 89.9% (+1.5%)
- Let's look at learnt weights:

well worth	13.788515
------------	-----------

best	13.633200
------	-----------

rare	13.570259
------	-----------

better than	13.500025
-------------	-----------

VS

bad	-24.467648
-----	------------

poor	-24.319746
------	------------

the worst	-23.773352
-----------	------------

waste	-22.880340
-------	------------

Near top positive

Near top negative