

Text classification – Amazon products reviews:

This task deals with text classification, one of the most common supervised tasks on text. A set of reviews in three diverse domains is attached to this assignment, and your goal is to predict a review's score (rating) based on its text. The ratings span values 1-5, meaning that is a 5-way classification.

Each review has the following format:

```
{"overall": 4.0, "verified": true, "reviewTime": "09 7, 2015", "reviewerID":  
"A2HDTDZZK1V4KD", "asin": "B0018CLO1I", "reviewerName": "Rachel E. Battaglia",  
"reviewText": "This is a great litter box! The door rarely gets stuck, it's  
easy to clean, and my cat likes it. I took one star off because the large is  
really small. My cat is 6 pounds and this isn't quite big enough for her. I'm  
ordering this same brand in a large. Online price is much cheaper than pets  
stores or other stores!", "summary": "Great Box, Get XL size!",  
"unixReviewTime": 1441584000}
```

where “overall” refers to the rating (the “label” you learn to predict), “review Text” to the body of the review, and “summary” to its summary.

Reviews are split into train (with 2000 reviews per class), and test (with 400 reviews per class) – you train a classifier on train data, evaluate and report the results on test data.

We don't have a validation set in this case since you are likely to work with classifiers' default parameters, i.e., no tuning is required.