

1. Amazon Fine Foods reviews

This dataset consists of reviews of fine foods from [amazon](#). The data span a period of more than 10 years, including all ~500,000 reviews up to October 2012. Reviews include product and user information, ratings, and a plaintext review. We also have reviews from all other Amazon categories.

Information about dataset: <http://snap.stanford.edu/data/web-FineFoods.html>

Download link: <http://snap.stanford.edu/data/finefoods.txt.gz>

Task description

Predicting Amazon product reviews' rating

Users can rate product on Amazon from 1 to 5 stars. The 5-star reviews contain many instances of "Great", "Good" and so on. On the other hand, the 1-star reviews use very little positive words and frequently contain warnings for potential customers. Create a model to predict the rating of the product based on the customer review.

Additional information

There are more than half a million reviews. Feel free to make a subset of the dataset to save your time.

Feel free to delete missing values.

Feel free to categorize rating to two features (good/bad rating)

The expected output contains source code, basic description of analysis, methods and your findings.

You should be prepared to answer questions about methods you used, process of the analysis and ideas to improve your predictive model.