Sentiment Analysis on Amazon Reviews of Digital Music

Edward Tong

Deep Learning project
NYC Data Science Academy

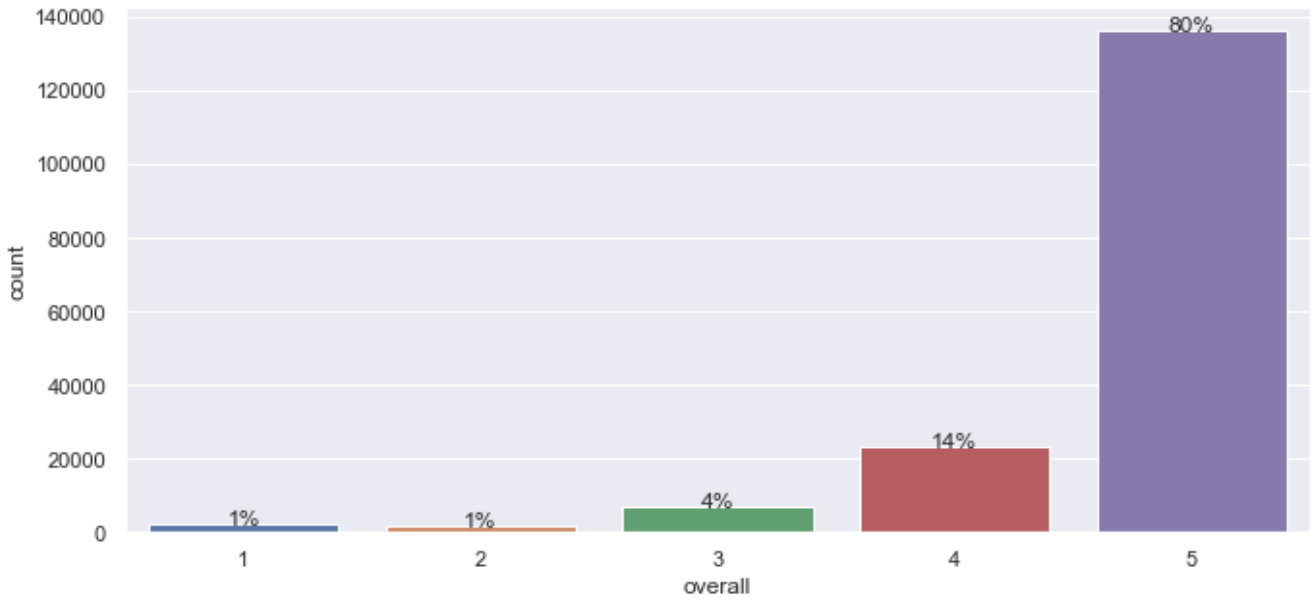
December 2019

Introduction

Predict binary ratings on Amazon Reviews dataset of digital music

- Data spans May 1996 - Oct 2018
- Sample of ∼170k reviews (subset of ∼1.5m total)
- 80:10:10 stratified split sampling
- ∼136k training, 17k testing, 17k validation
- Review Ratings binary label - Low (1, 2 stars) vs High (3, 4, 5 stars)

Distribution of ratings



Model approaches

Four classification model architectures considered

- `lightgbm` + `tf-idf` + Bayesian Optimization
- `lightgbm` + `fastText`, skip-gram + Bayesian Optimization
- `Embedding-LSTM-Sigmoid`
- `Embedding-LSTM-Dropout-Dense(128-Relu)-Dropout-Dense(64-Relu)-Dropout-Sigmoid`

Model performance | Out-of-sample

Model Architecture	**AUC** Validation set
`lightgbm + tf-idf + bayes-optimization`	0.941
`lightgbm + fastText + bayes-optimization`	0.935
`Embedding-LSTM-Sigmoid`	0.897
`Embedding-LSTM-Dropout-Dense(128-Relu)-Dropout-Dense(64-Relu)-Dropout-Sigmoid`	0.903

Discussion

- `lightgbm` + `tf-idf` has highest AUC, `fastText` is competitive
- Character level *n*-grams performed better than word level equivalent
- `fastText` offers insightful cosine similarity between words
- LSTM results are inconclusive due to lack of in-depth tuning, more time required for model development, e.g.
  - `fastText` embedding should be considered as input to LSTM