

# Naive Bayes Analysis of Amazon Reviews

Chris Brown and AJ Walters

# Dataset

Other researchers scraped 35 million Amazon reviews from an 18 year period

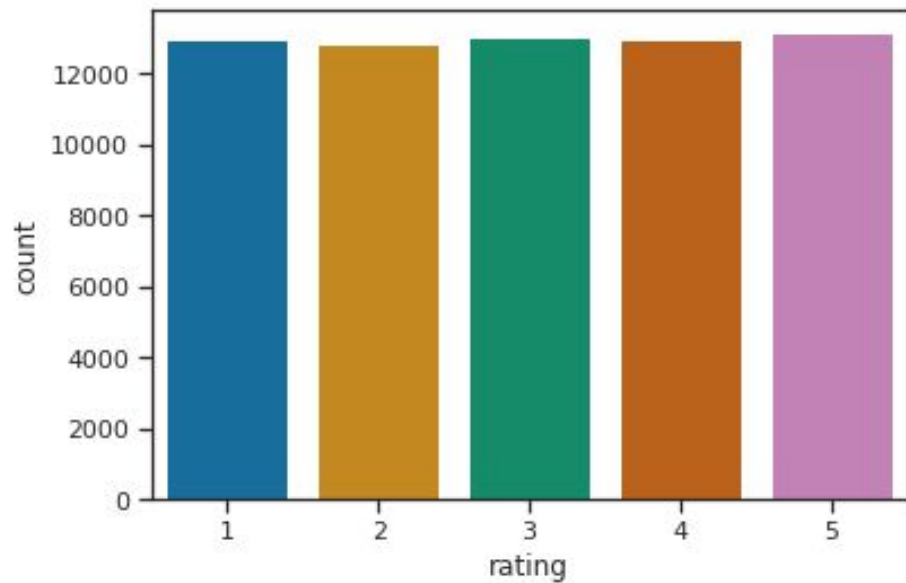
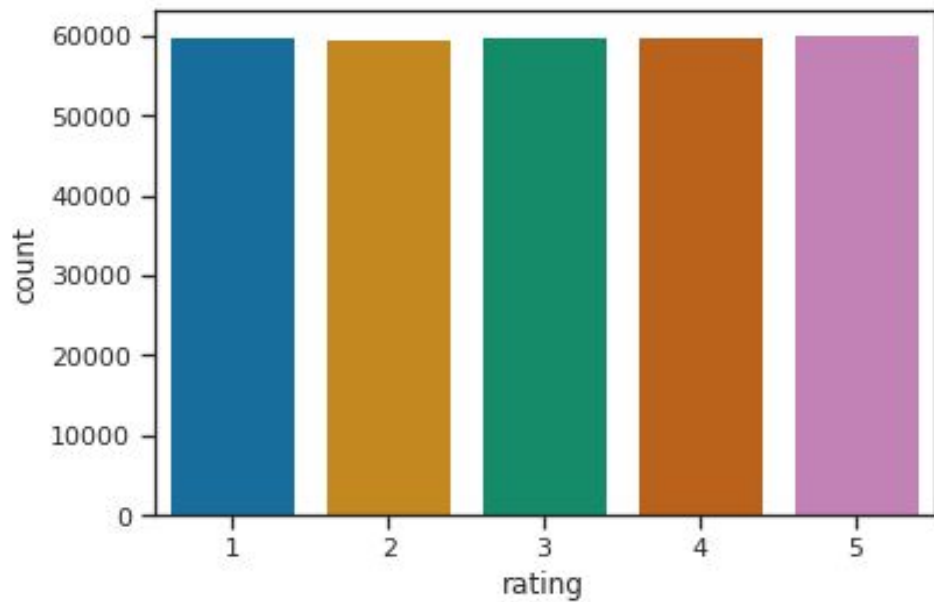
Has rating, title, and review text

Dataset randomly sampled into a CSV: 3 million training, 650,000 testing

Even distribution of ratings

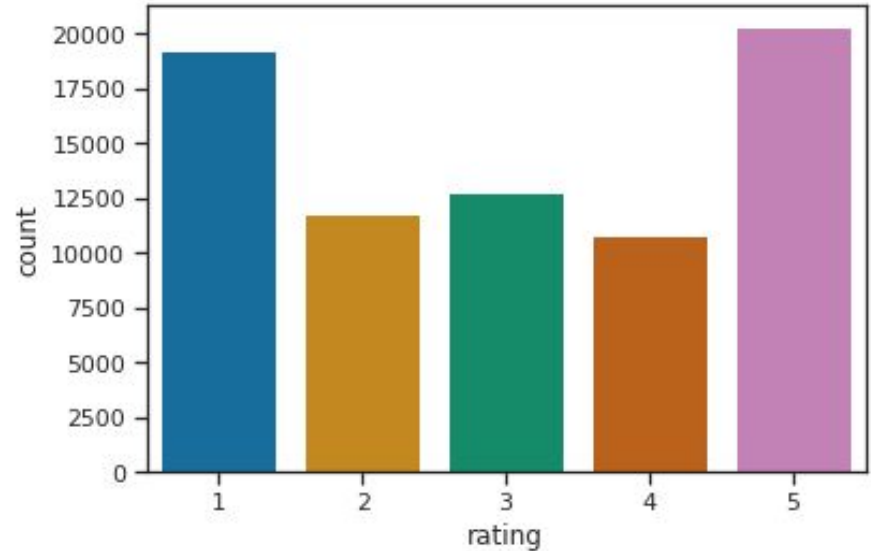
We reduced that dataset by to 10% (300,000 training, 65,000 testing)

# Dataset



# Bernoulli Naive Bayes

| Bernoulli | Precision | Recall | F1-Score | Support |
|-----------|-----------|--------|----------|---------|
| 1         | 0.52      | 0.66   | 0.58     | 15075   |
| 2         | 0.41      | 0.33   | 0.37     | 14852   |
| 3         | 0.40      | 0.34   | 0.37     | 14789   |
| 4         | 0.43      | 0.30   | 0.35     | 15215   |
| 5         | 0.50      | 0.67   | 0.57     | 15068   |



# Multinomial Naive Bayes

| Multinomial | Precision | Recall | F1-Score | Support |
|-------------|-----------|--------|----------|---------|
| 1           | 0.53      | 0.63   | 0.57     | 15075   |
| 2           | 0.40      | 0.37   | 0.39     | 14852   |
| 3           | 0.38      | 0.35   | 0.36     | 14789   |
| 4           | 0.41      | 0.34   | 0.37     | 15215   |
| 5           | 0.55      | 0.60   | 0.58     | 15068   |

