On Building a Response Prioritization Engine for Southwest Air

Debashri Mukherjee[1], Kathryn Tooker[2], Tyreke White[3]

CS 229, Autumn 2017, Stanford University

Motivation:

Responding to disgruntled customers is a top business priority that should be streamlined to maximize customer satisfaction. Social media has enabled customers to both voice their grievances and share their praise directly with airlines in real time. Replying to tweets is a daunting task that can be augmented by machine learning. We address this need for Southwest Airlines by applying sentiment analysis to tweets and predicting customer service agent response times to inform tweet response prioritization.

Data:

- Sourced data by using Tweepy, a Python library for accessing Twitter API
- Initial dataset contained over 50,000 tweets (~Nov 15-Dec 10).
- After implementing matching criteria, workable dataset was approximately 1,000 tweets

Methods:

Sentiment Analysis

- Labeling: VADER (vs. original sentiment dictionary)
- Classification Techniques Applied:
  - ❖ Naive Bayes
  - ❖ Support Vector Machines (SVM)
  - ❖ Neural Networks (NN)
  - ❖ Non-negative Matrix Factorization (NMF) Topic Modeling

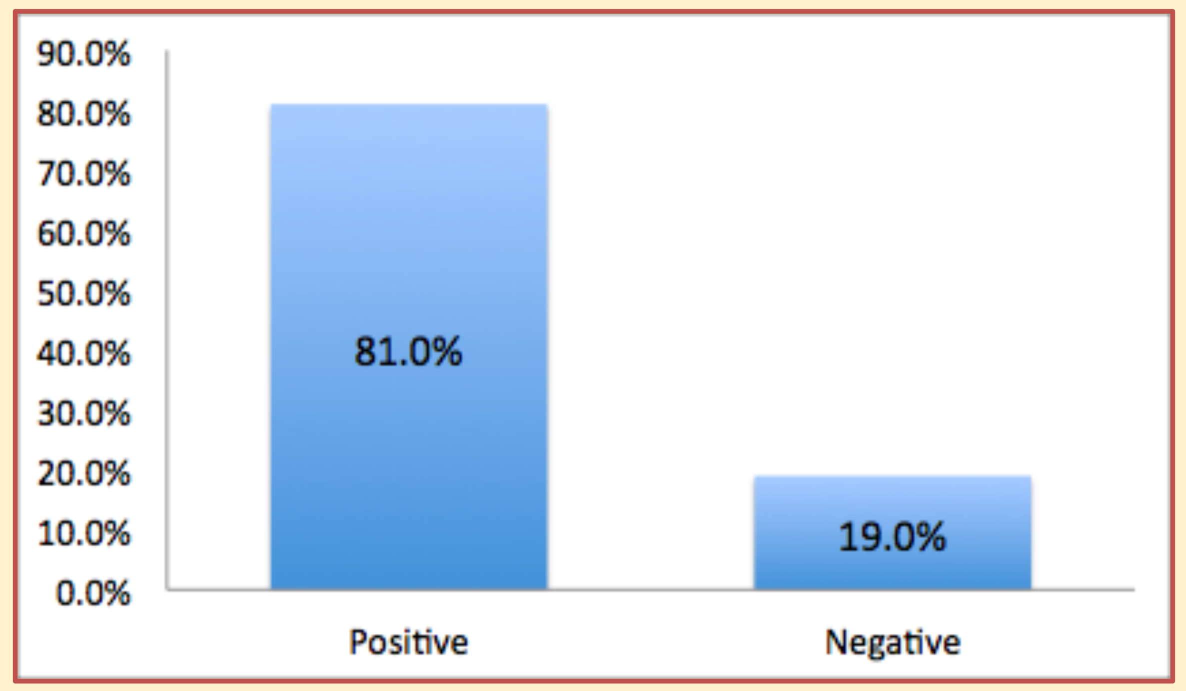
Response Time Prediction

- Ordinary Least Squares Regression
- Lasso Regularization
- Ridge Regression
- PLS (Partial Least Squares) Regression with PCA

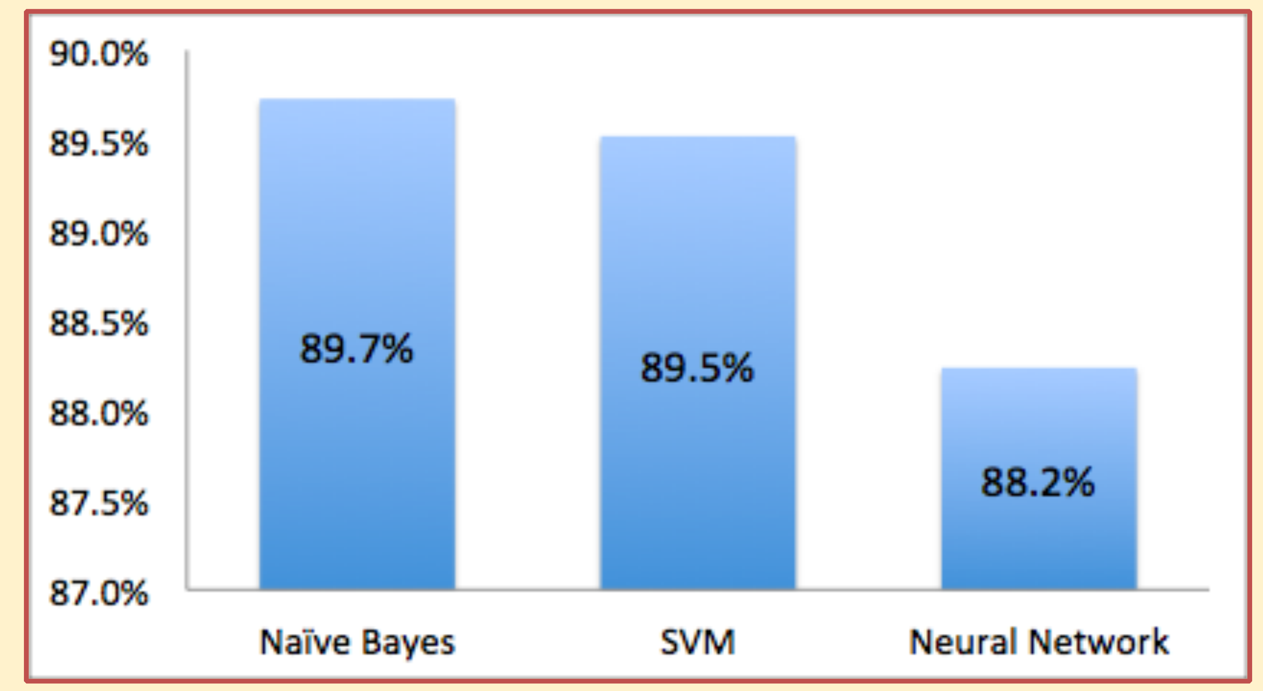
Results:

Sentiment Analysis

Distribution of Sentiment Labels



Classification F1 Scores



"@SouthwestAir Was a long wait, but as usual, a wonderful agent quickly resolved problems! #lovesouthwest"

"@SouthwestAir My sister has been on hold for an hour and we leave at 7am tomorrow! This is outrageous"

NMF Topics

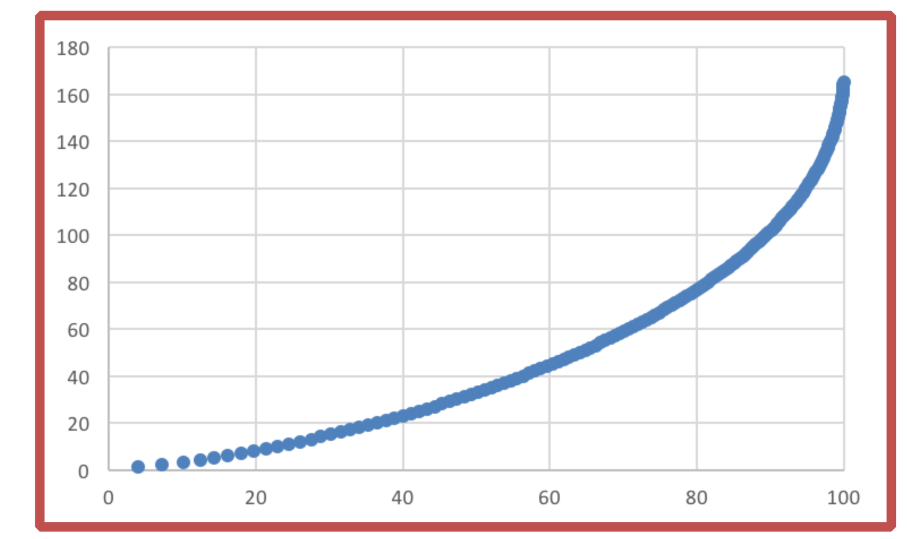
1. amp great getting help thanks → **Customer Service**
2. southwest flying good help thank → **Customer Service**
3. crew btw late great flight →**In-flight Experience**
4. southwest swa know guys love → **Customer Relationship**
5. southwest getting time holidays happy → **Positive Brand Image**

Response Time Prediction

Regression R-squared Scores

Lasso	Ridge Regression	PLS with PCA
0.051	0.133	0.540

Variance explained by principal components



PLS, even with the principal components that explain 100% of the variance in the dataset, gave an R-squared score of slightly more than 0.5.

Discussion:

Sentiment Analysis

- Initially attempted 3-5 labels, but models performed poorly due to incorrect labeling. Ultimately pursued binary classification (strong positive/negative).
- Many more positive than negative labels.
- NMF topics are not very distinct. This may already be too narrow a scope of study, evidenced by high frequency of repeated words (e.g. help, flight, thanks).
- **Future Work:**
  - More accurate labeling, particularly for neutral tweets.

Response Time Prediction

- Lasso and Ridge regression did not work on our data, likely because they did not appropriately capture the high multicollinearity in the data.
- PCA, however, performed better because it captured more feature information rather than removing/reducing coefficients.
- Future Work:
  - Current model looks biased - plan to include the exogenous features, other than the current independent variables, that could be affecting the response time.

[1] debm@stanford.edu, [2] ktooker@stanford.edu, [3] ttwhite@stanford.edu