

US Airlines Service Recommendation

Tianshu Ji, Wanyi Qian, and Xiaotong Duan
 Departments of Chemical Engineering, Stanford University
 Stanford, California



Abstract

The goal of this project is to 1) generate a model for six major U.S. airlines, including United Airlines, Delta, Southwest, that performs sentiment analysis on customer reviews so that the airlines can have fast and concise result on their performances, 2) make recommendations on what are the most important aspect of their services they could improve given customers' complains. In this project, we performed multi-class classification using Naive Bayes, SVM and Neural Network on the Twitter US Airline data set obtained from Kaggle.

DATA SET

1). Sentiment Analysis

Sentiment	Airline	Text
Positive	Virgin America	@ Virgin America it was amazing, and arrived an hour early. You're too good to me.
negative	United	@united flight arrives 30 minutes early, but then have we to wait for an hour for our bags.
neutral	Delta	At the airport ready to get this @JetBlue red eye going.... Soooooo sleepy. #NoPlaceLikeHome #eventhoughits2degreesathome

2). Negative Reason Classification

Reason	Airline	Text
Bad Flight	United	@united Lovely new plane from LGA to ORD but no power outlets?
Cancelled Flight	Southwest	@SouthwestAir can't believe how many paying customers you left high and dry with no reason for flight Cancelled Flightlations Monday out of BDL! Wow.
Customer Service Issue	United	@united ok it's now been 7 months waiting to hear from airline. I gave them quite a bit more than the 30 days requested! Terrible service
Damaged Luggage	United	@united when will I hear? Guitar was damaged in December. I use my guitar to earn a living. Get your act together!
Flight Attendant Complaints	United	@united have an employee at the gate 15min before boarding like u expect ur customers to. Be a competent company like ur rivals
Long lines	US Airways	@USAirways hundreds of people in line and less than half the desks being manned at CLT. Help?
Lost Luggage	Southwest	@SouthwestAir 2 hrs to put a tag on my bag sayin it should go to greenville instead of Raleigh?! ARE YOU KIDDING ME?!

3). Data Distribution

Around **60% negative**, **20% positive**, and **20% neutral** twitters

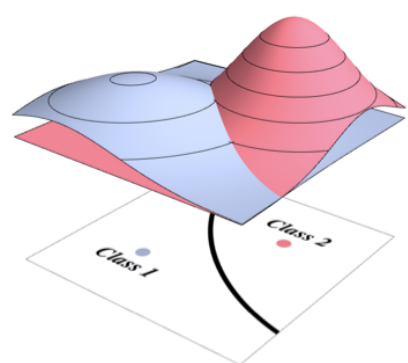
The total data set is composed of **3.5% Virgin Airlines**, **19% American Airlines**, **20% US Airways**, **15% Delta**, **26% United**, and **16.5% Southwest**.

MODEL

Naïve Bayes ^[1]

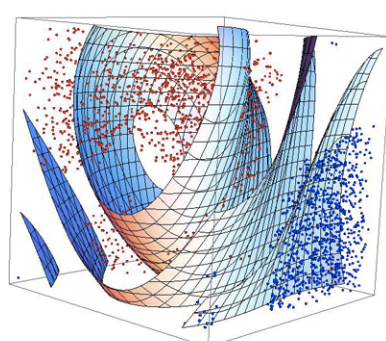
Implement Naïve Bayes Classifier using package sklearn

$$p(C_k|\mathbf{x}) = \frac{p(C_k) p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$



Support Vector Machine ^[2]

Implement Support Vector Machine using the package sklearn. Having tuning different Kernel function, we found the best test accuracy using linear Kernel, LinearSVC.



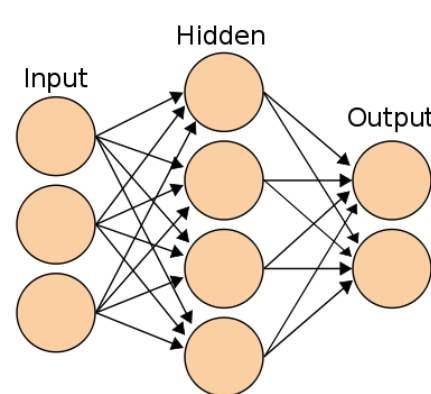
One Layer Neural Network ^[3]

Implement the following model by Tensorflow, with learning rate = 0.01

$$h = Wx + b$$

$$y_{\hat{}} = \text{softmax}(h)$$

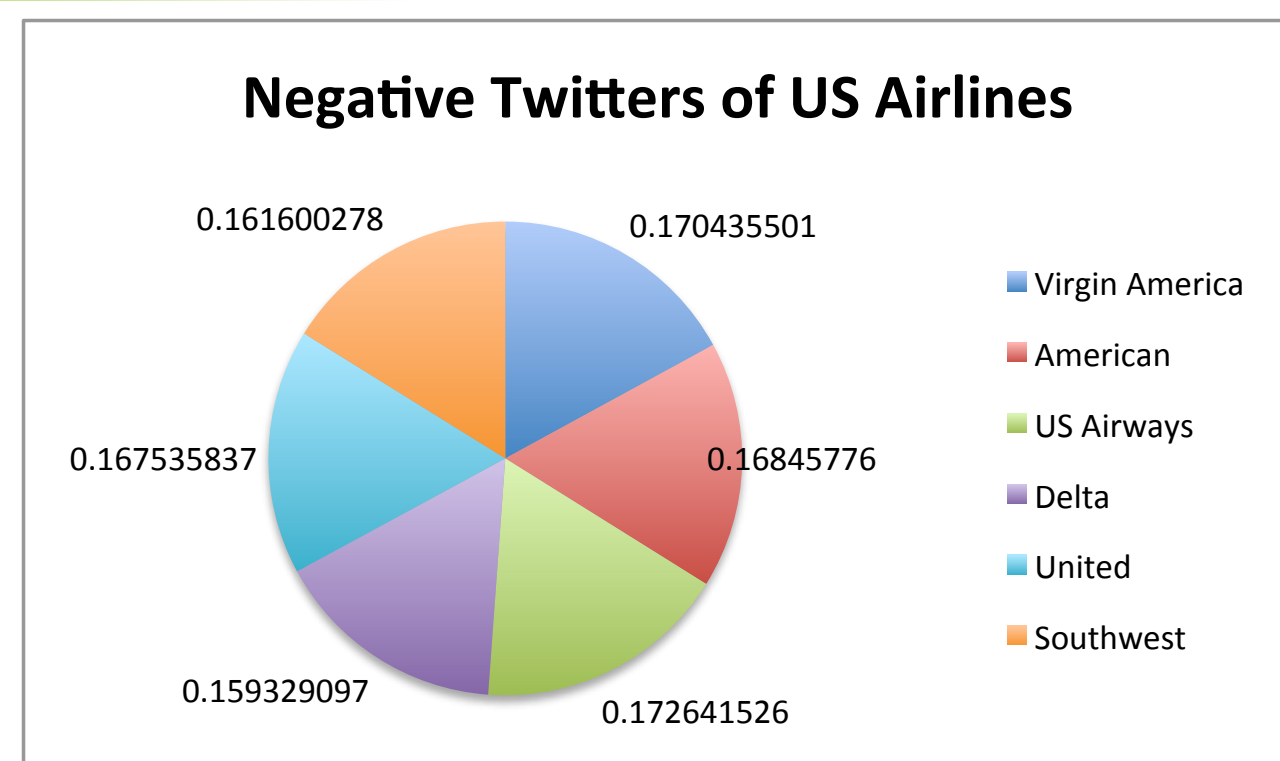
$$J = \text{CE}(y, y_{\hat{}})$$



	Model	Test Error
Sentiment Analysis	Naïve Bayes	0.41 0.38
	SVM	0.46 0.46
	One layer neural network	0.2633
Negative Reason Classification	Naïve Bayes	0.368
	SVM	0.389
	One layer neural network	0.3781

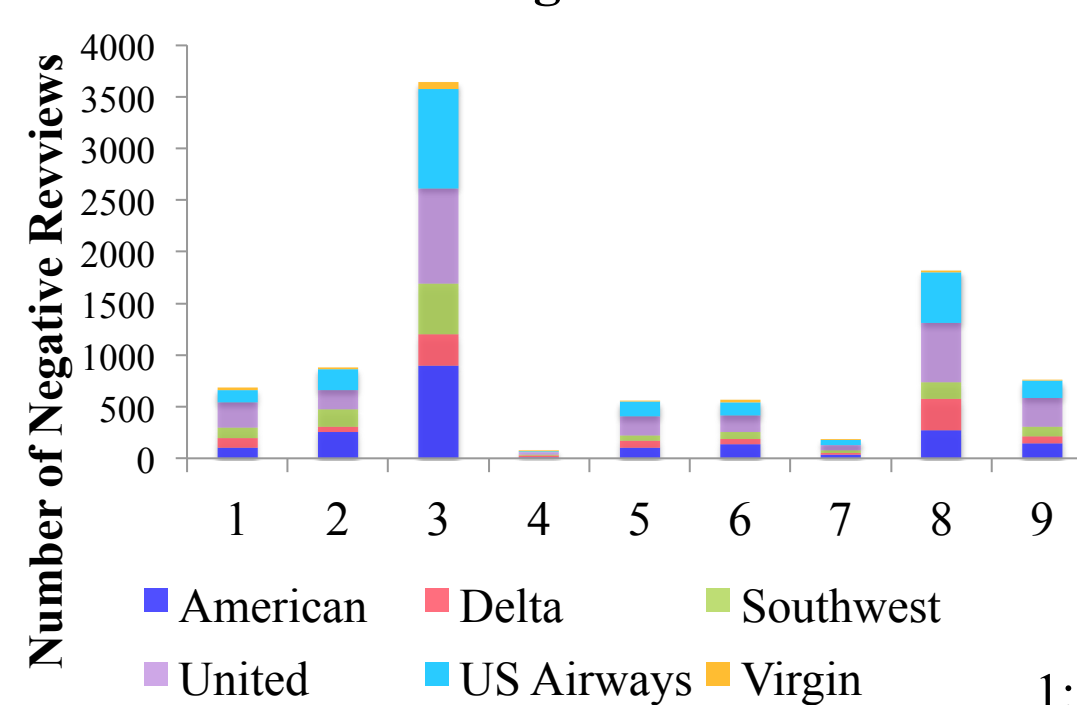
RESULT

Sentiment Analysis

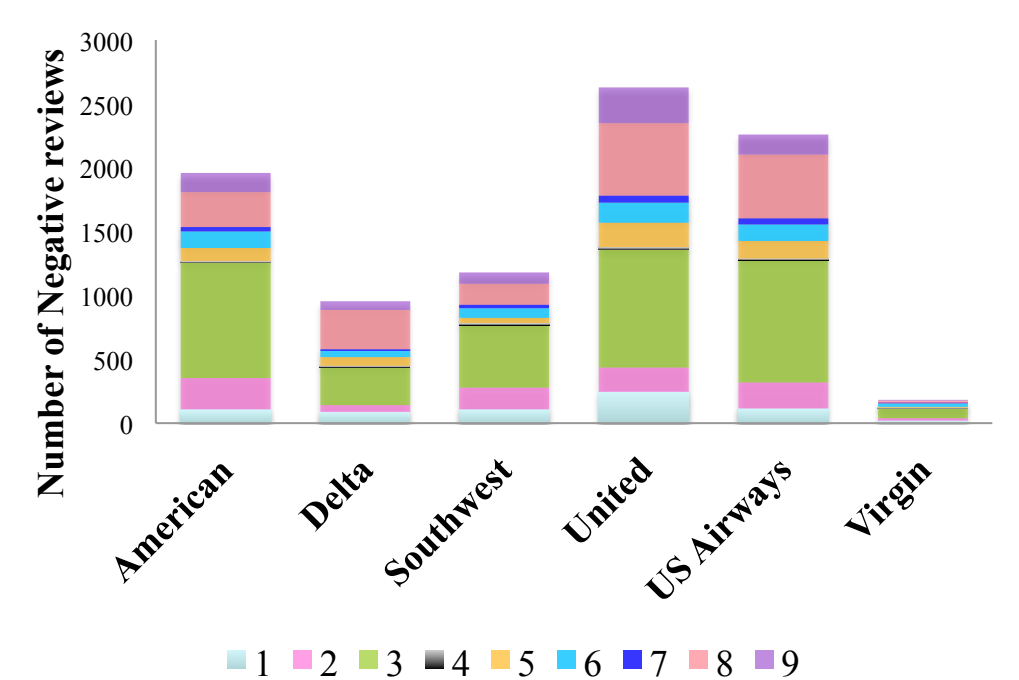


Recommendation for Airlines

Based on Negative Reason



Based on Airlines



- 1: Bad Flight
- 2: Cancelled Flight
- 3: Customer Service Issues
- 4: Damaged Luggage
- 5: Flight Attendant Problems
- 6: Booking Problems
- 7: Long Lines
- 8: Late flight
- 9: Lost Luggage

KEY TAKEAWAYS

- ◆ In sentiment Analysis part, **one layer neural network** is recommended according to its **74%** high accuracy compared with SVM and Naïve Bayes
- ◆ In negative reason classification, **Naïve Bayes** is recommended according to its lowest test error, giving **63.2%** accuracy.
- ◆ **Customer service issue** is the **most popular** negative reason and **Delta** is the **most competitive** airline according to its number of negative reviews and distribution of negative reason.