



STEVENS
INSTITUTE *of* TECHNOLOGY
THE INNOVATION UNIVERSITY®

Airline Sentiment Analysis Using Twitter Data

BIA 658 - Social Network Analysis
11/27/2017

Instructor : Feng Mai
Group2 : Yuyi Yan, Fanshu Li, Biyan Xie



1 Introduction

Content

2 Data Collection

3 Data Analysis

4 Application & Discussion



Introduction

Airlines have been increasingly used in today's society.

There is an increasing number of people heading to social media such as Facebook and Twitter to talk to airline companies.

People share their traveling experience and express some positive or negative comments.





Project Goals

- Proportion of negative sentiment tweets
- Reasons for negative sentiment
- Location negative sentiment analysis
- Clustering analysis of words
- Build prediction model based on text



1 Introduction

Content

2 Data Collection

3 Data Analysis

4 Application & Discussion



Data Collection

Southwest[®] 

American Airlines 

**United
Airlines** 


DELTA
AIR LINES

 **america**


U·S AIRWAYS



Data Resource



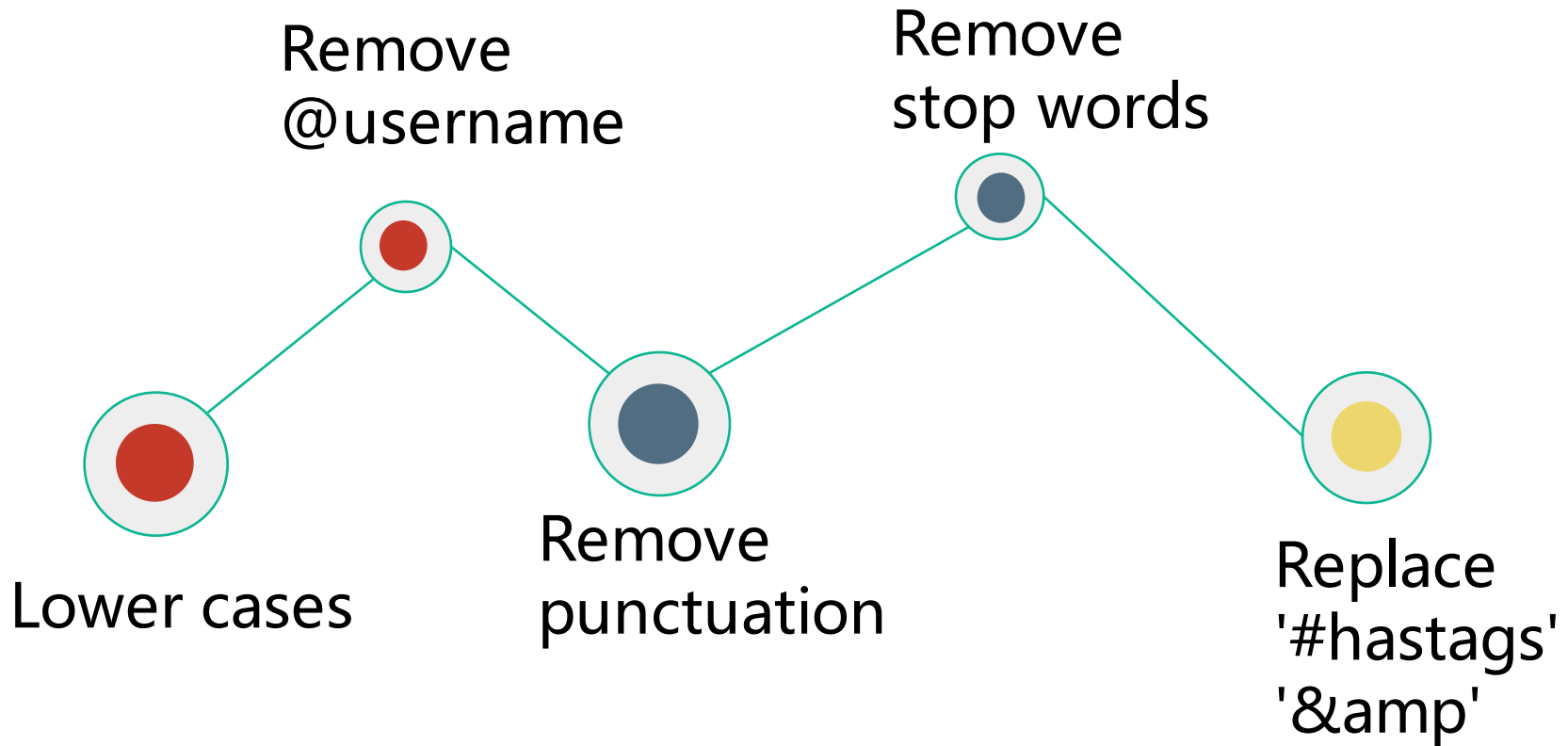
- *Crowdflower's Data for Everyone library*
7700 users
Scraped from February of 2015

```
> dim(tweet)
[1] 9079  15
```

```
> names(tweets)
[1] "tweet_id"           "airline_sentiment"      "airline_sentiment_confidence" "negativereason"
[5] "negativereason_confidence" "airline"                "airline_sentiment_gold"      "name"
[9] "negativereason_gold"      "retweet_count"          "text"                        "tweet_coord"
[13] "tweet_created"           "tweet_location"         "user_timezone"
```



Data Pre-processing





Method

1. Data visualization

2. Prediction model



1 Introduction

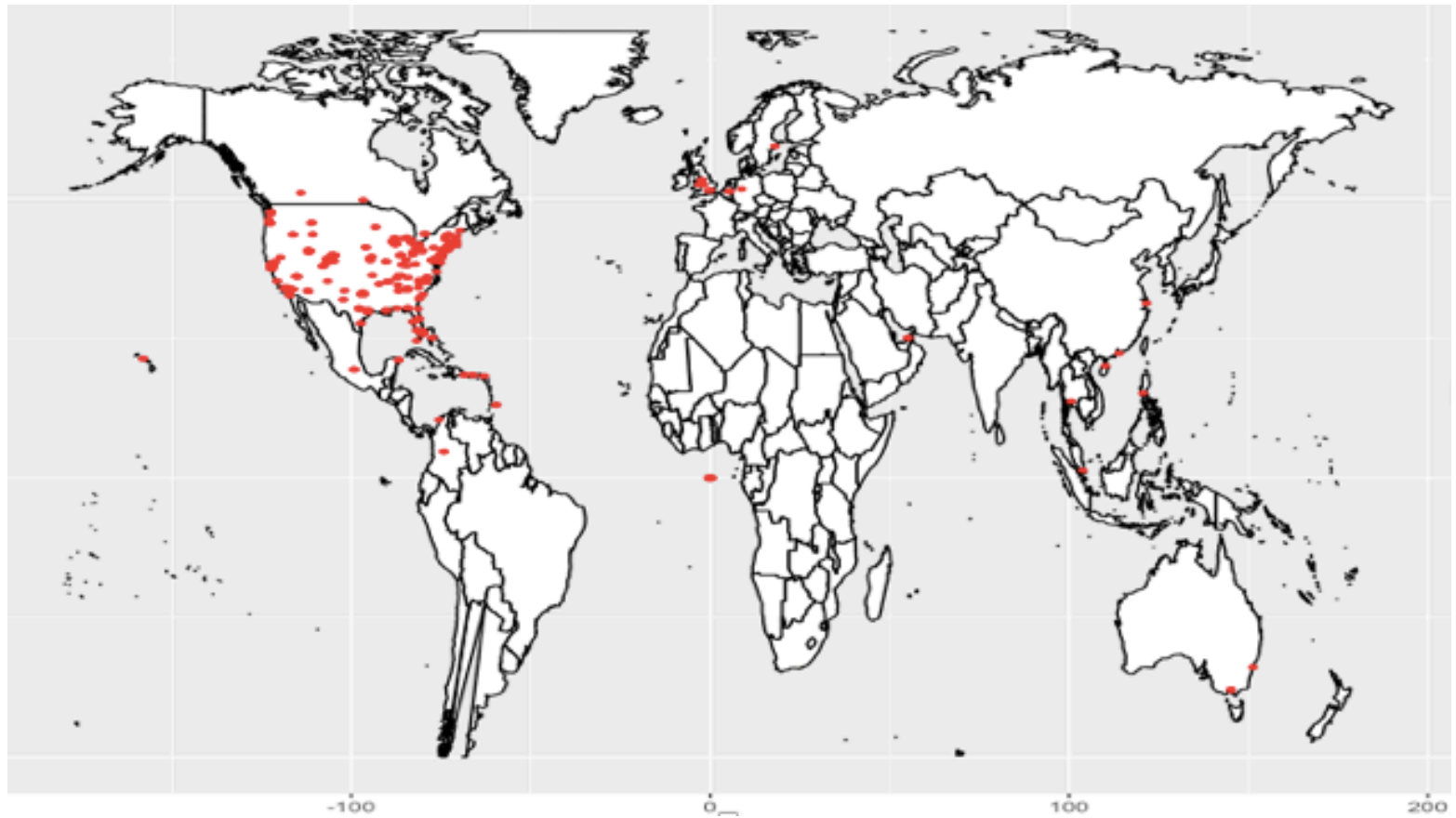
Content

2 Data Collection

3 Data Analysis

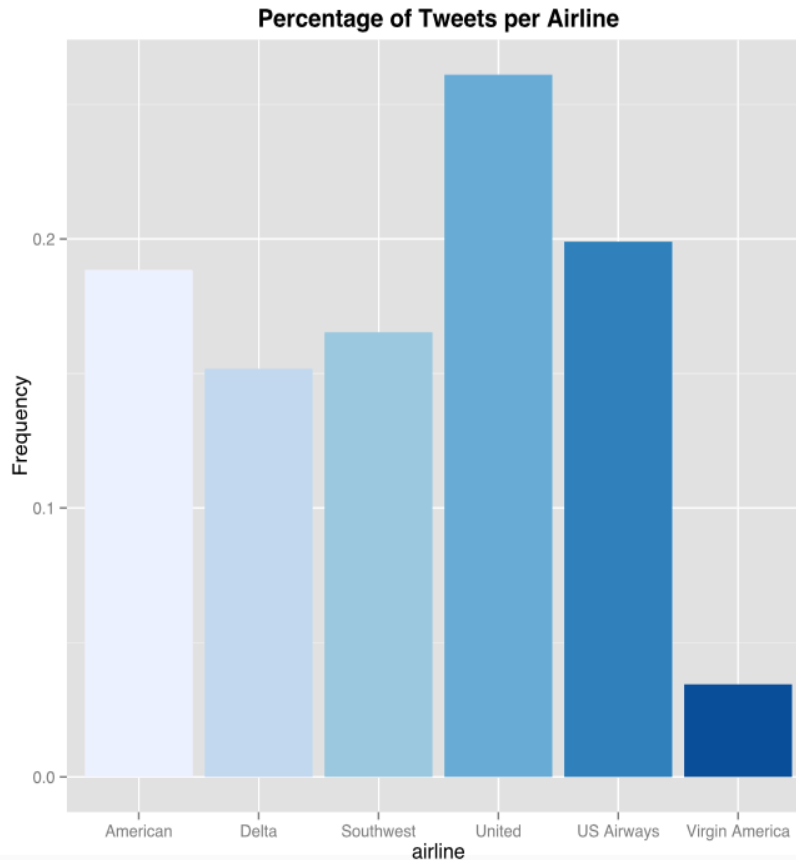
4 Application & Discussion

The locations of tweets across the world





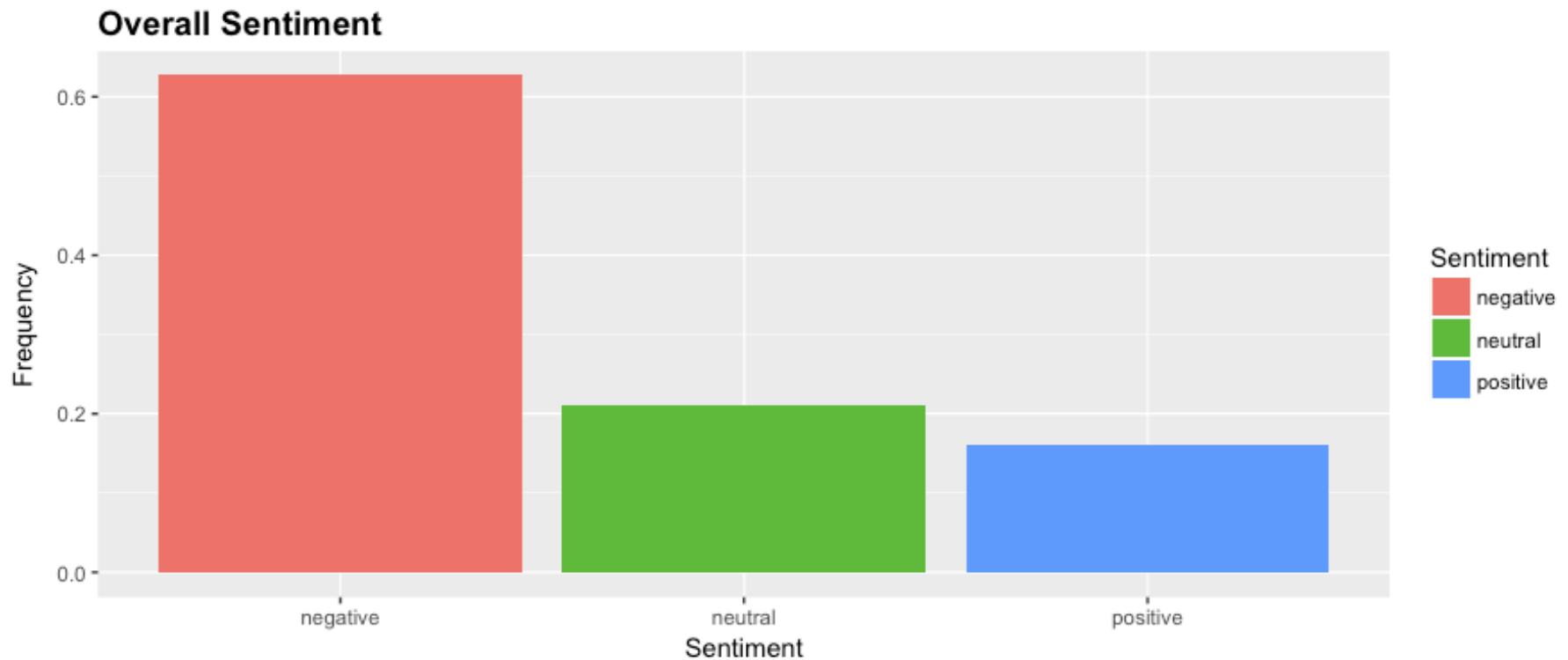
The proportion of tweets per airline



American	Delta	Southwest	United
0.18845628	0.15177596	0.16530055	0.26106557
US Airways	Virgin America		
0.19897541	0.03442623		

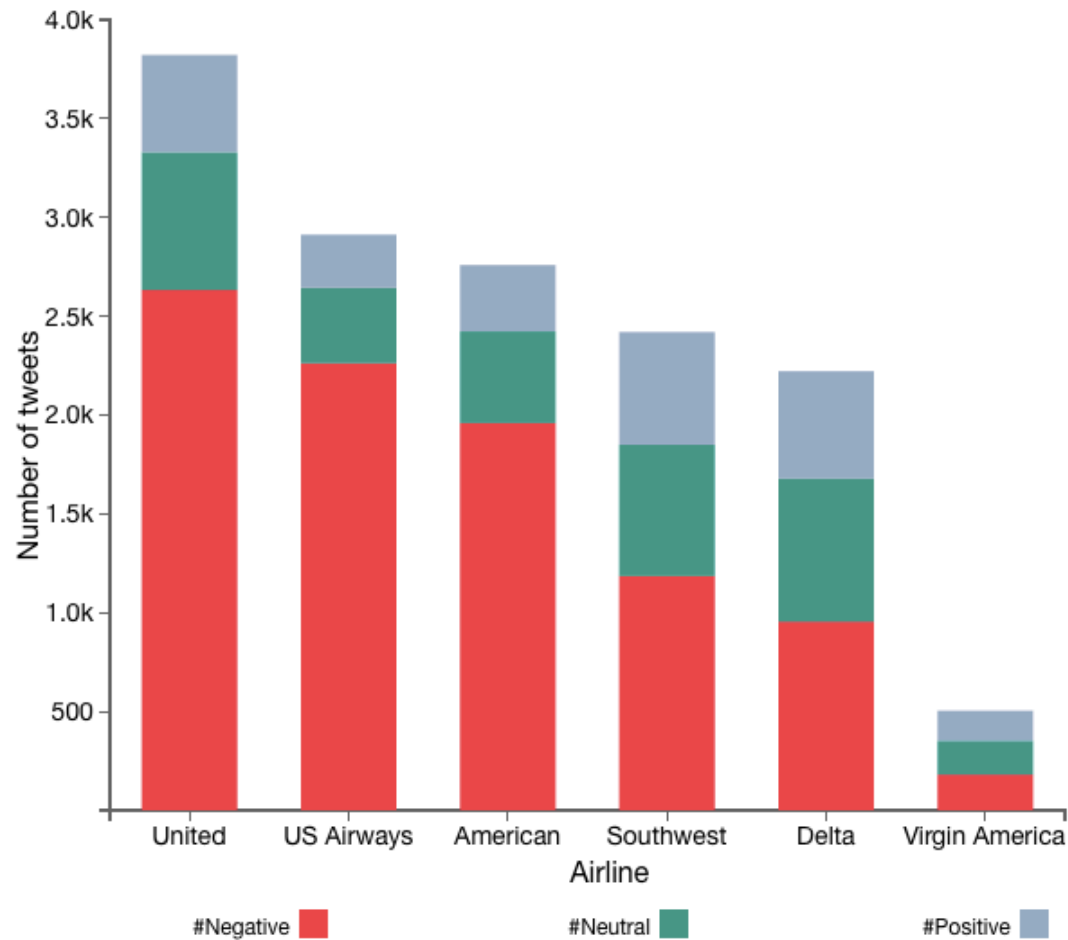


The proportion of negative sentiment



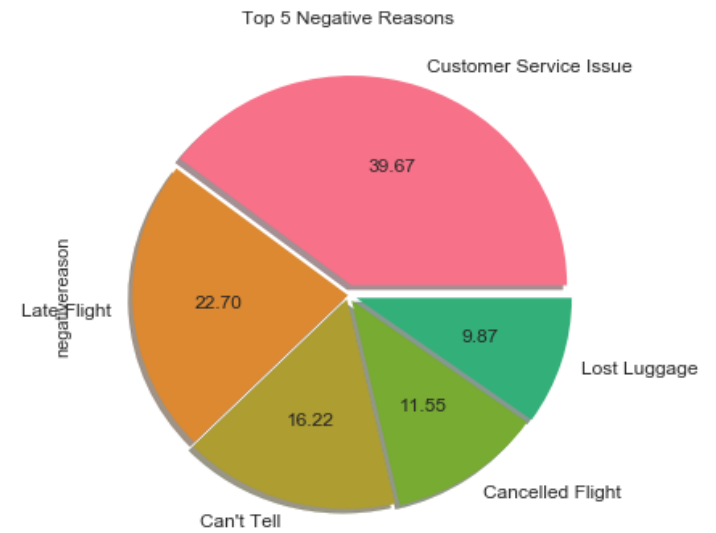
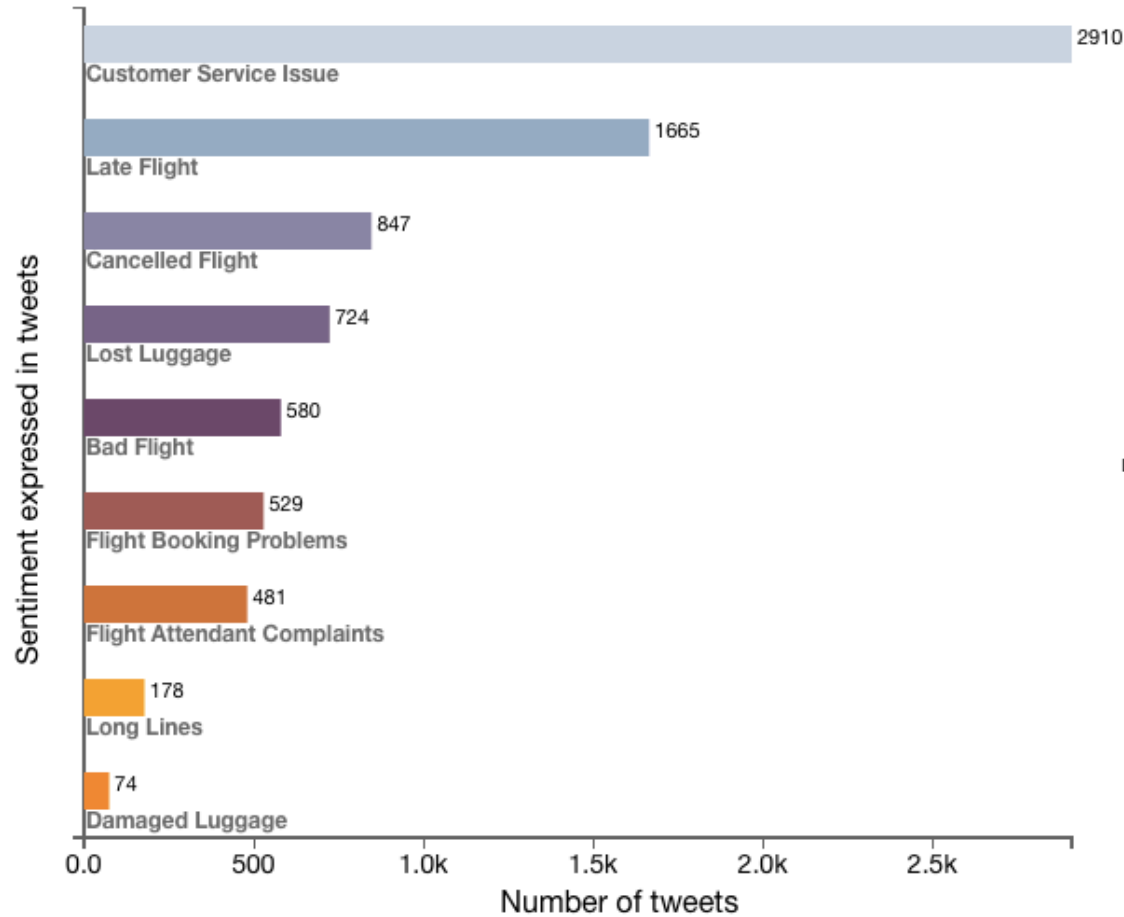


The proportion of negative sentiment Tweets per airline



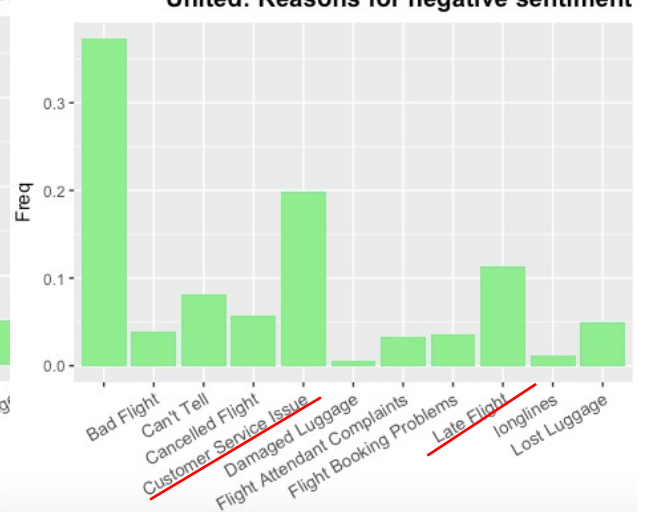
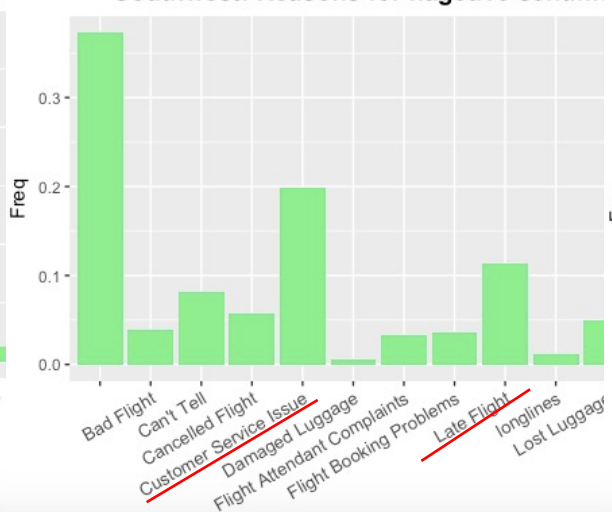
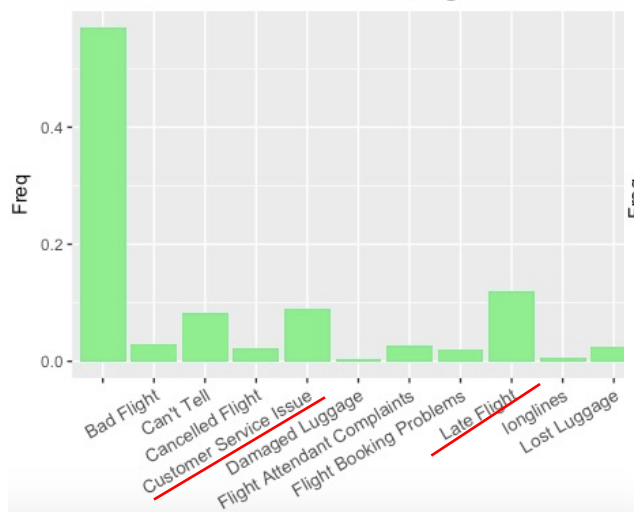
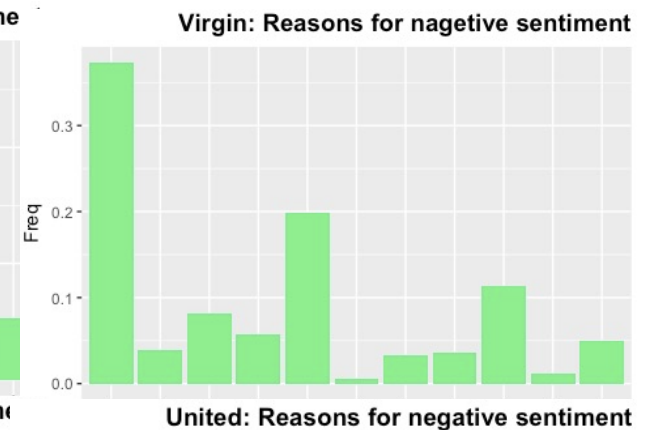
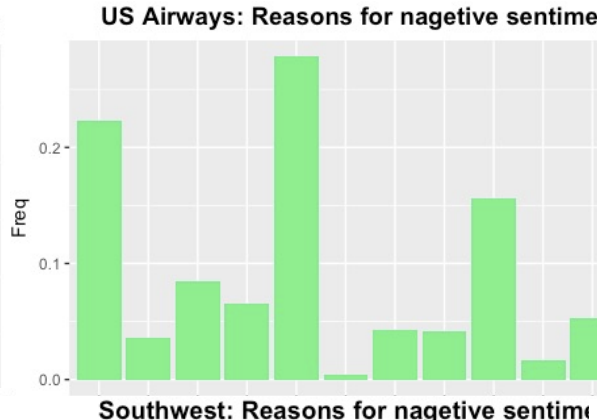
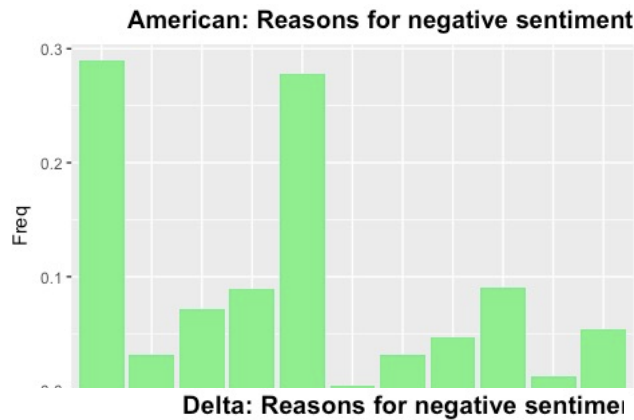


Reasons for negative comments

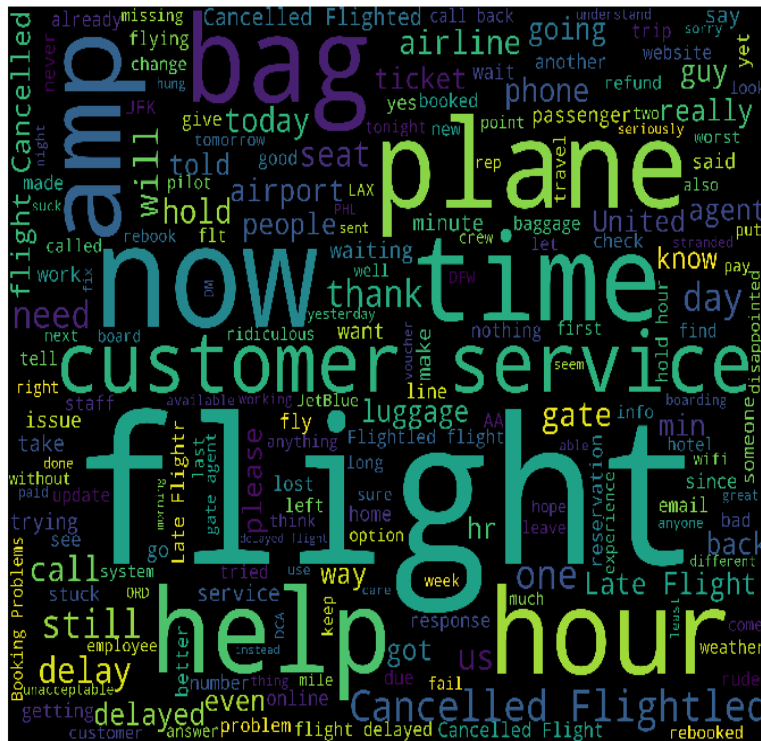




Normalized reasons for negative sentiment per airline



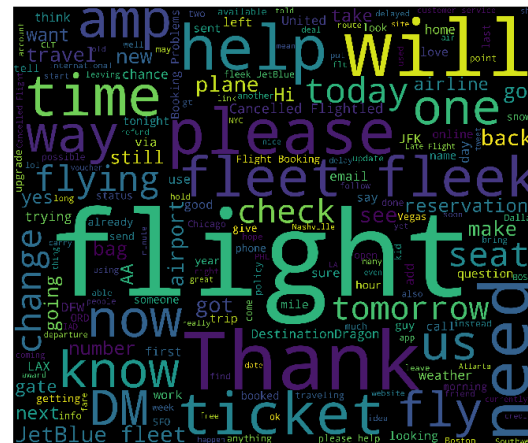
Wordclouds of Twitter comments



Negative



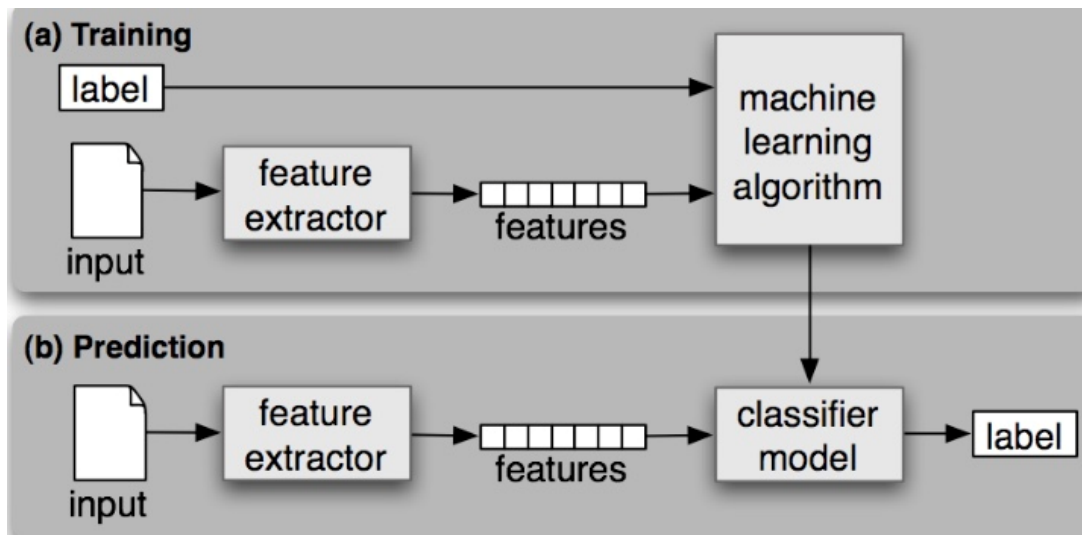
Positive



Neutral



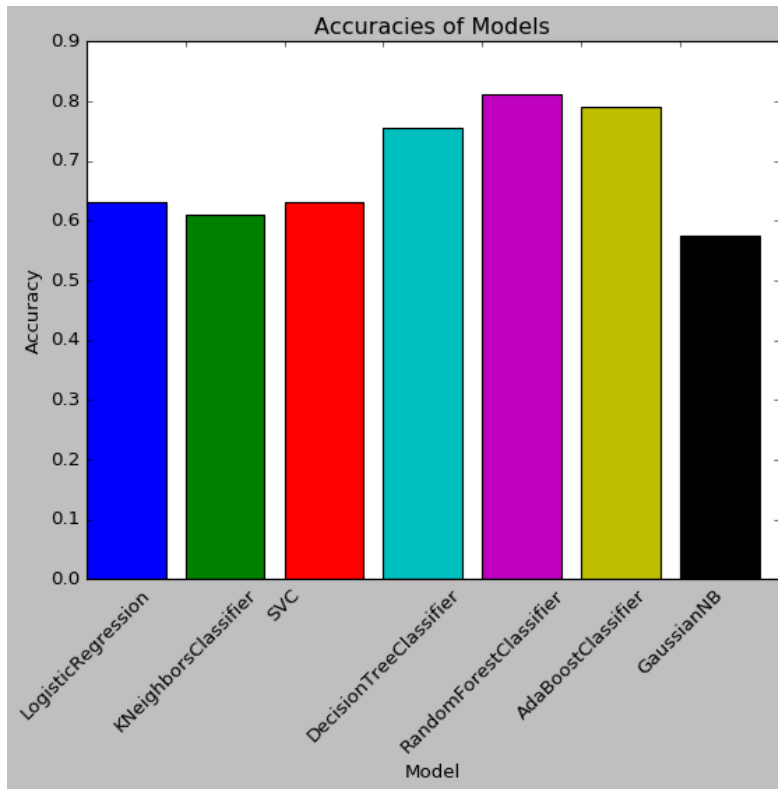
Prediction Models



- Logistic Regression
- K-nearest neighbors
- Support Vector (SVC)
- Decision Tree
- Random Forest
- AdaBoost
- GaussianNB



Accuracy of Models



Comparison of accuracy

Random Forest : 0.810792349727

AdaBoost : 0.790300546448

Decision Tree : 0.755464480874

Logistic Regression: 0.631147540984

SVC : 0.631147540984

KNN: 0.609289617486

GaussianNB : 0.575136612022



1 Introduction

Content

2 Data Collection

3 Data Analysis

4 Application & Discussion



Application

Customer service, late flight, cancelled flight are the major negative reasons for all airlines.

- Reply to any comments as soon as possible and obtain detailed reasons.

The flight experience is terrible!@AmericanAir

Thank you for your attention!
Which service you are not satisfied with?

Effective flight
service !@AmericanAir

Thank you for your attention! We make
greater efforts with your supports.



Application

- Airline companies can have a specific Twitter page for customer service in order to obtain more detailed customer feedback.





Application

- Assign to different departments based on the classification model

My Luggage is
lost! @JetBlue



Luggage Department



Discussion

1. Twitter sentiment only reveal limited feedback
2. Latest data is difficult to obtain
3. Clustering analysis to discover words association
4. Better Model performance
- 5.....





STEVENS
INSTITUTE *of* TECHNOLOGY
THE INNOVATION UNIVERSITY®

Thanks for listening.

