

# SENTIMENT CLASSIFICATION OF TWITTER DATA: AMERICAN AIRLINES CUSTOMER SERVICE

Amir Ghaderi

Mezbah Uddin

Md. Shariful Islam

MSc Data Science and Analytics

# Overview



- Problem and Goal
- Data Preprocessing
- Sentiment analysis
- Text analysis
- Classification
- Network analysis
- Conclusions

# Problem and Goal



## □ Problem

- ▣ Airlines using twitter for Customer Service reasons
- ▣ How to extract relevant information from large amounts of tweets

## □ Goal

- ▣ Develop a system for airline companies to process mass amounts of twitter data in an streamlined approach

# Data Preprocessing

- Twitter Search API
- Keywords{"@Delta", "@WestJet"}
- Three Visualizations
  - ▣ Distribution of tweet length
  - ▣ Number of Unique Users
  - ▣ Retweet Percentage



# Sentiment Analysis: SentiSrength



Positive  $\geq 2$  and  
Negative = -1



Positive

Positive  $\geq 2$  and  
Negative = -1



Negative

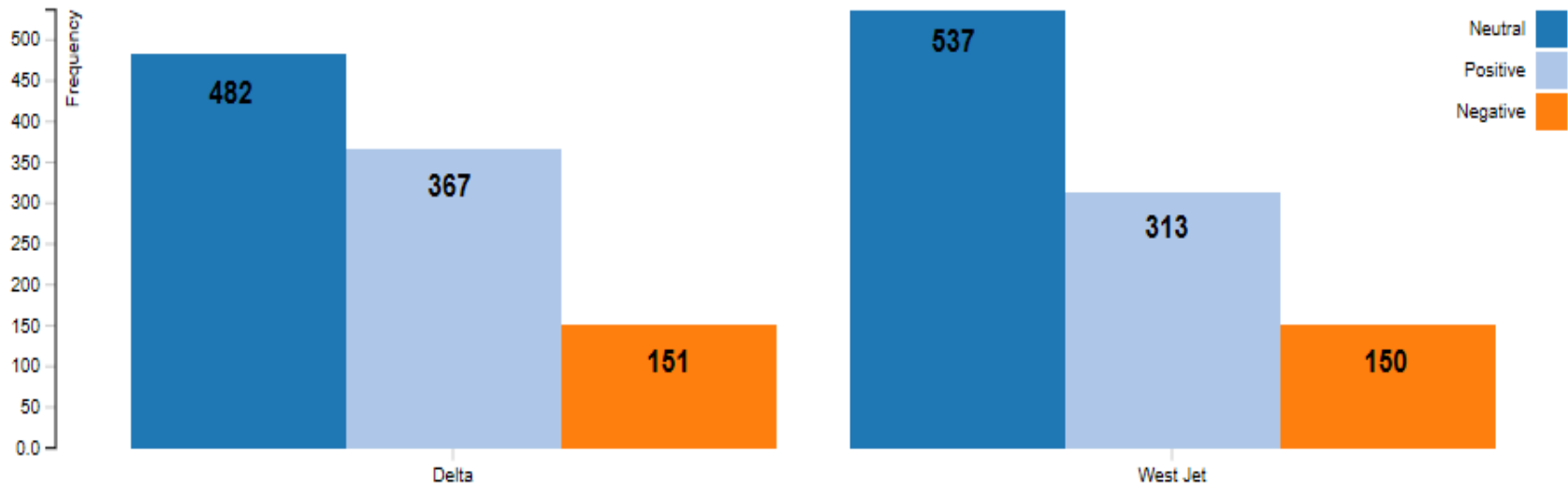
Not Positive and Not  
Negative



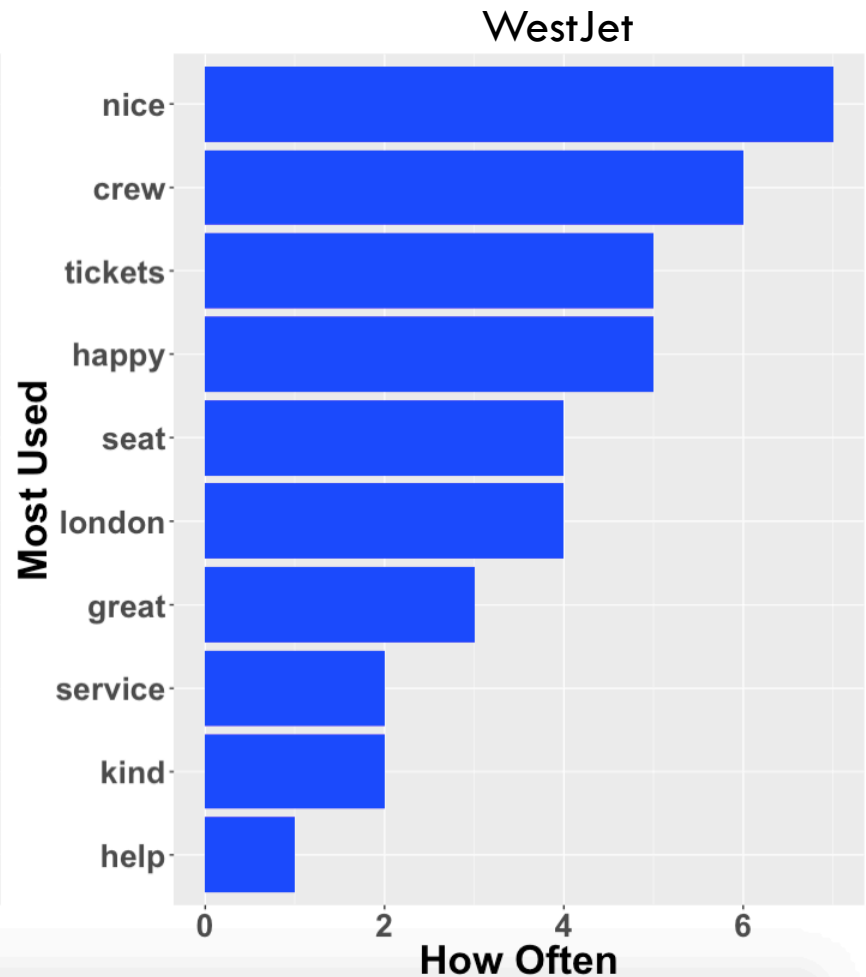
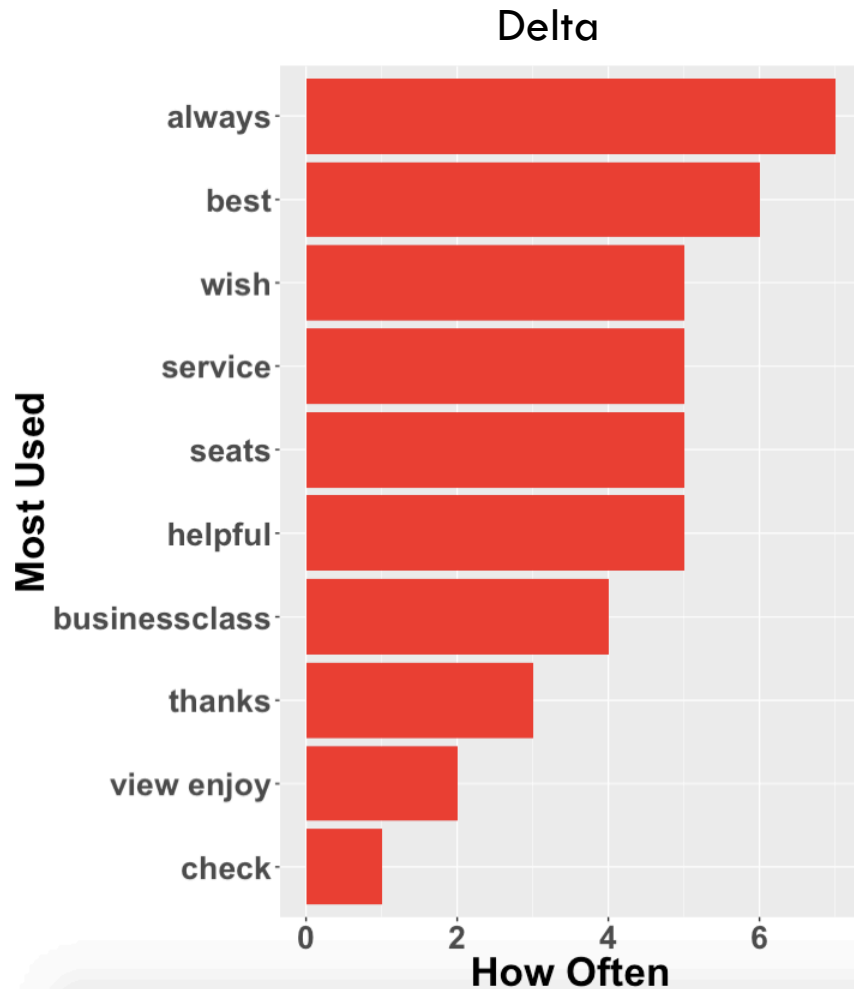
Neutral

# Sentiment Analysis

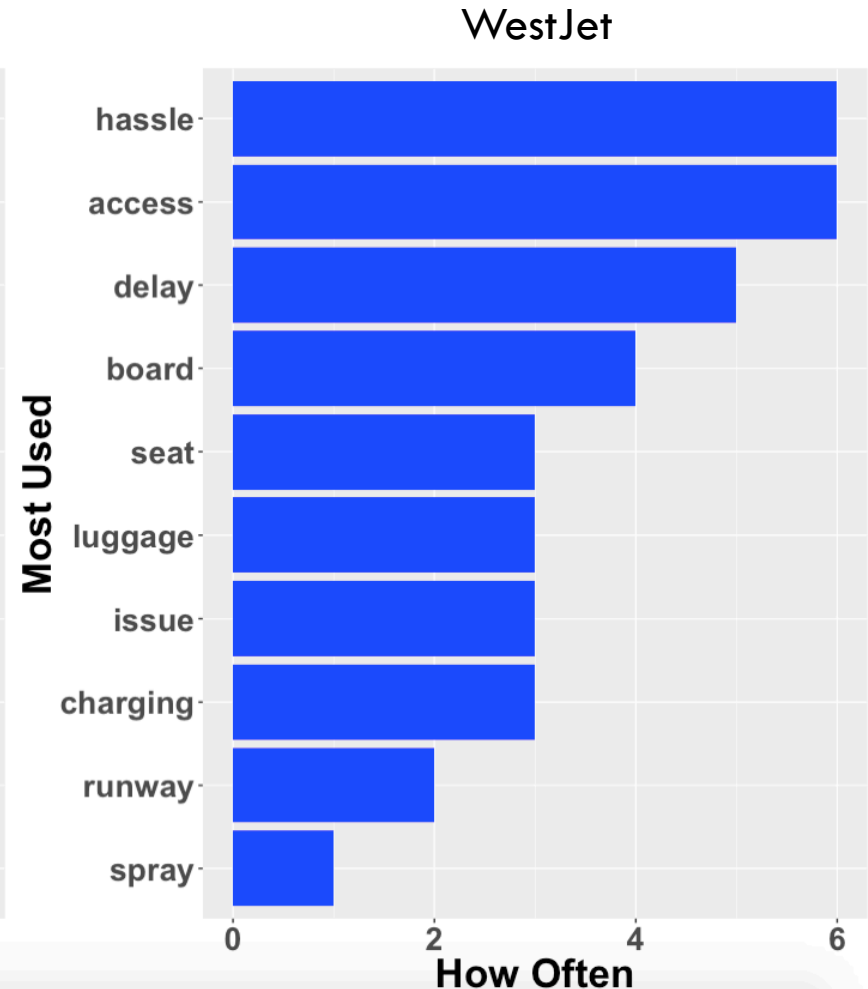
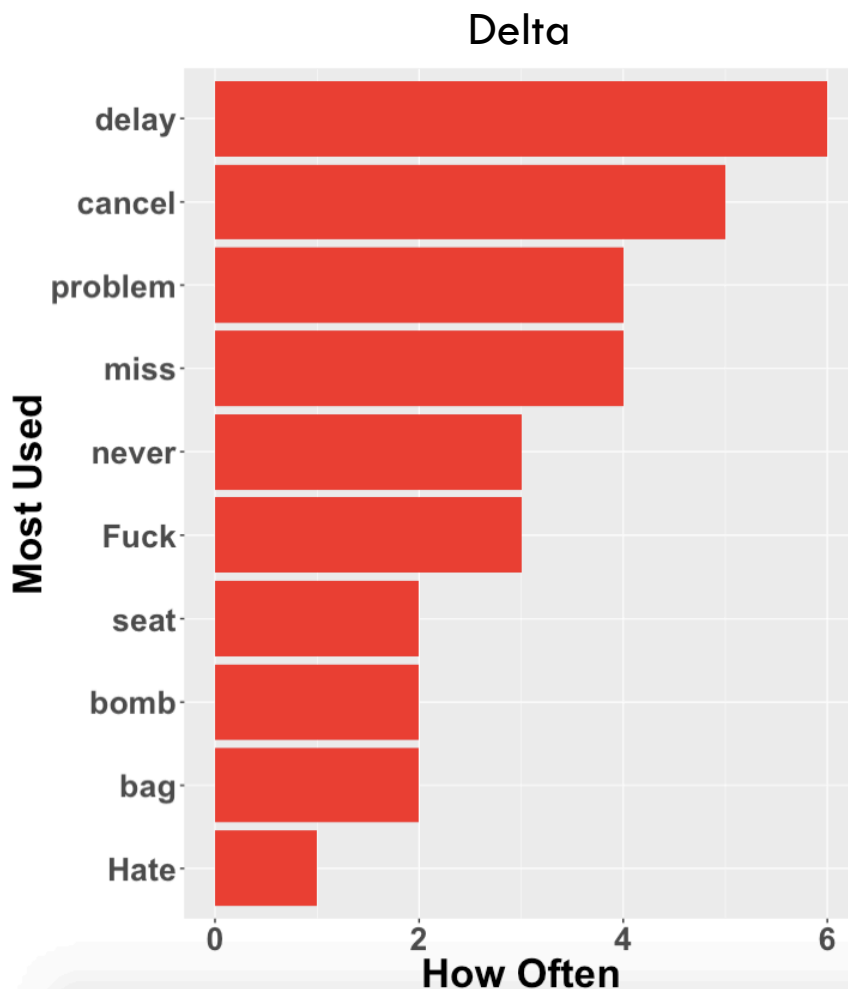
## Sentiment Analysis of Airline Tweets



# Text Analysis: Positive words



# Text Analysis: Negative words





# Classification

- Classifiers{Max Entropy, Tree Method, SVM, Random Forest, Bagging}
- 4 Fold Cross Validation
- Delta
  - ▣ Training size - 400 Tweets
  - ▣ Testing size - 127
- WestJet
  - ▣ Training size - 363 Tweets
  - ▣ Testing size - 100

# Classification: Delta

	Fold 1	Fold 2	Fold 3	Fold 4	Average
Max Entropy	100%	100%	100%	100%	100%
Tree Method	77%	74%	79%	73%	73%
SVM	88%	86%	80%	85%	85.1%
Random Forest	84%	76%	85%	73%	79%
Bagging	85%	82%	78%	81%	82%



# Classification: WestJet

	Fold 1	Fold 2	Fold 3	Fold 4	Average
Max Entropy	95%	91%	93%	90%	92%
Tree Method	55%	58%	65%	64%	64%
SVM	68%	58%	65%	64%	85.1%
Random Forest	68%	68%	73%	62%	68%
Bagging	65%	61%	65%	63%	63%



# Network analysis



## Network Properties: ?

Diameter: 9  
Density: 0.002311908173666  
Reciprocity: 0.023733455043359  
Centralization: 0.409113978658343  
Modularity: 0.422835263616733



## Network Properties: ?

Diameter: 14  
Density: 0.001752374832977  
Reciprocity: 0.026388888888889  
Centralization: 0.395366677643213  
Modularity: 0.486988811292125

# Conclusion

---

- Social Media Viable Platform For Customer Service
- Developed a system to process tweets
  - ▣ Sentiment analysis {positive, Negative, Neutral}
  - ▣ Text analytics {Relevant Word frequencies}
  - ▣ Machine Learning Classifiers
  - ▣ Network Analysis

discover  
questions  
where? how? why asking questions challenge who?  
clues  
**QUESTIONS**

ask  
who?  
discover

what?

when? knowing  
investigation

clues  
how

why?  
ask

knowing  
investigation