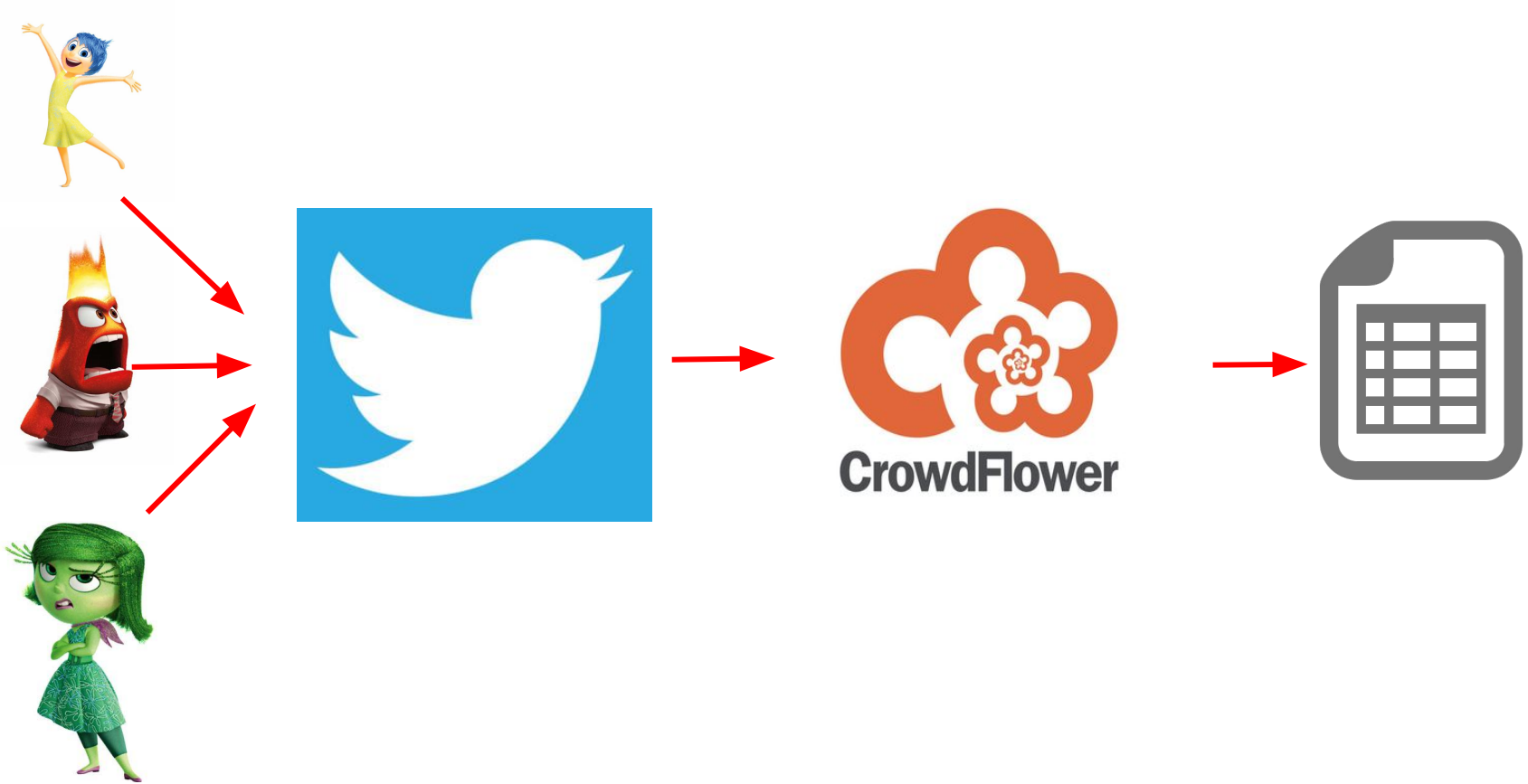


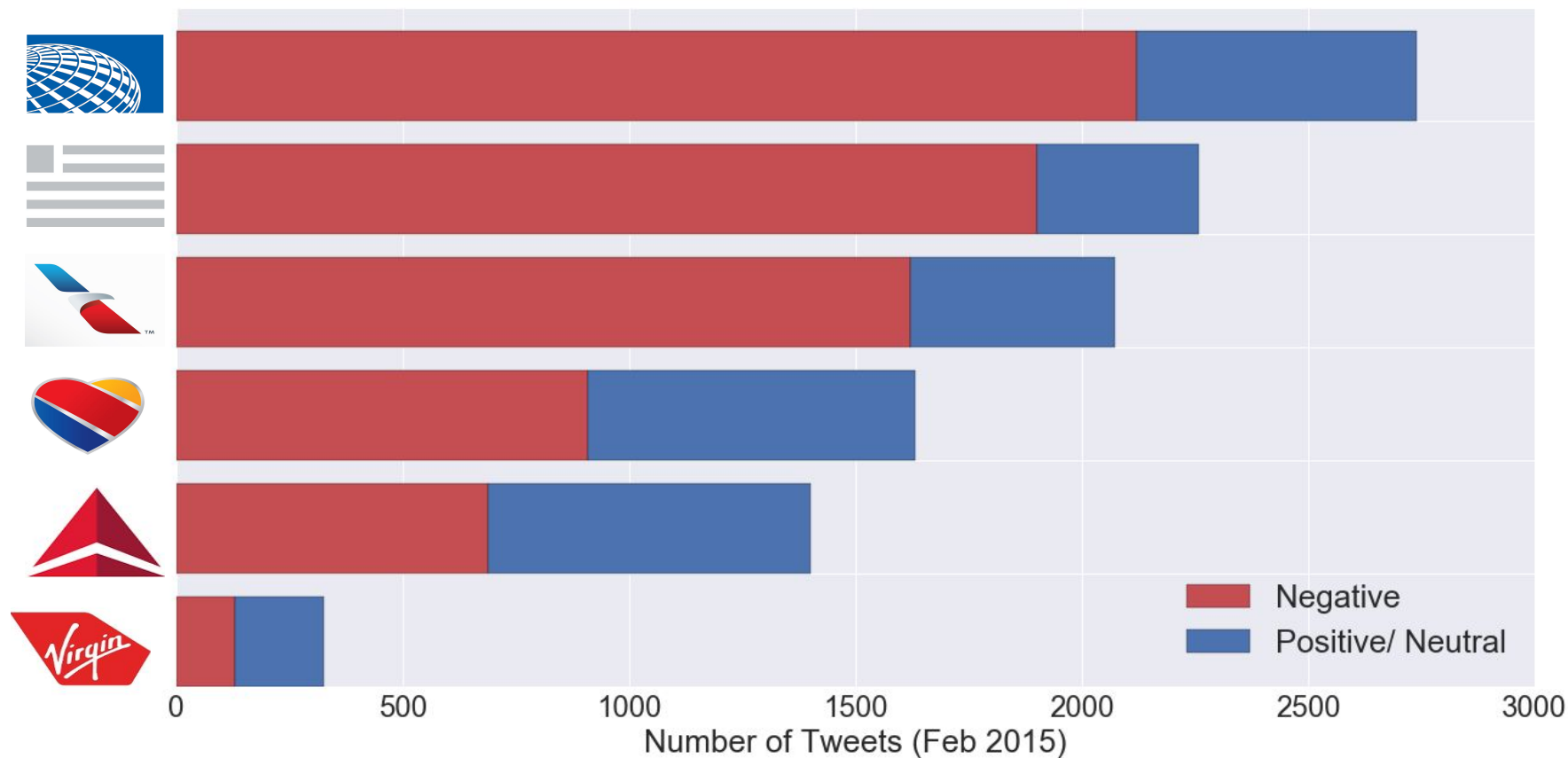


# Fly the Frustrating Skies: Sentiment Analysis of the U.S. Airline Industry

# Data Pipeline



# Tweets by Sentiment and Airline

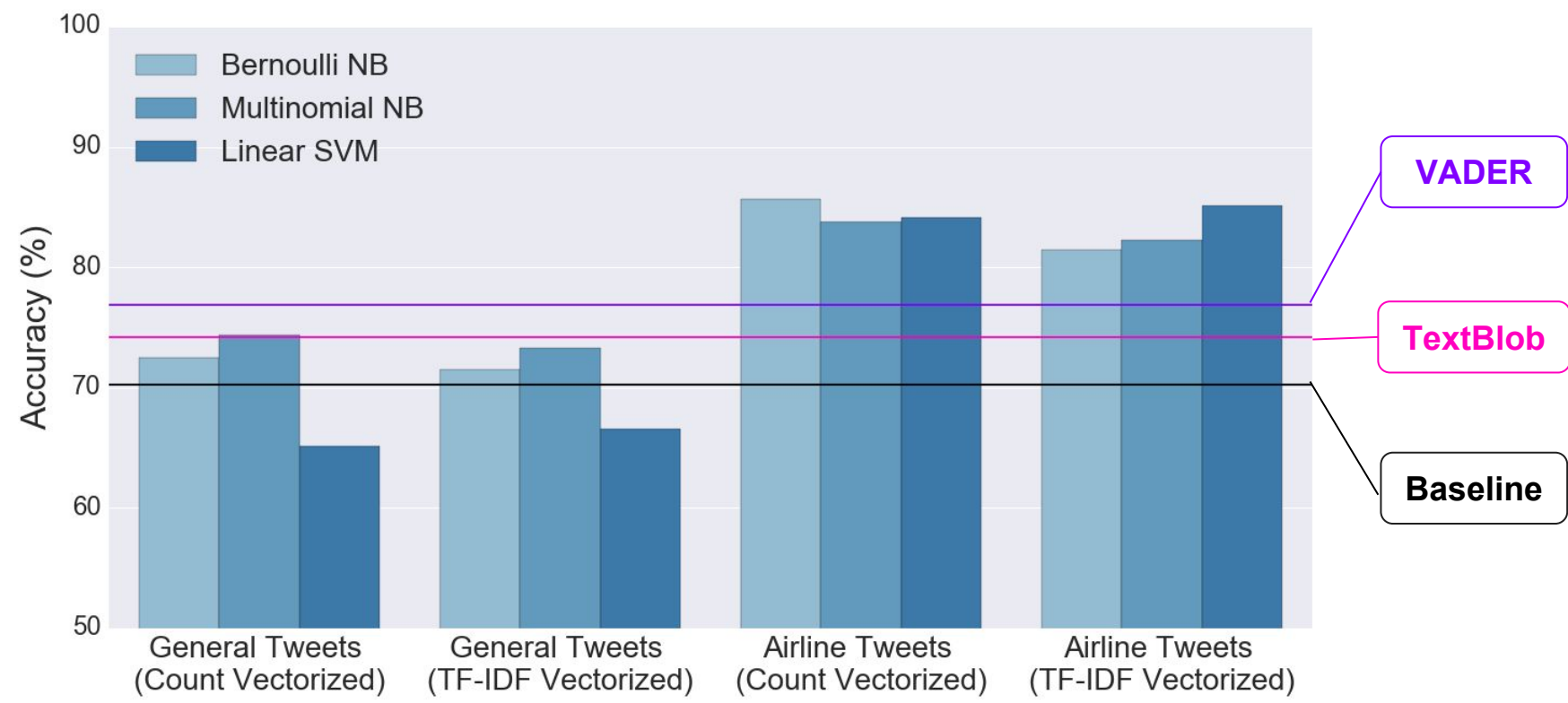


# Questions

1. How does the performance of different sentiment analyzers compare?
  - a. Lexicon-Based
    - i. TextBlob
    - ii. VADER (Valence Aware Dictionary and Sentiment Reasoner)
  - b. Machine-Learning Based (Naive Bayes & SVM)
    - i. General Tweet Trained
    - ii. Airline-Specific Tweet Trained
  
2. Which tweet keywords are most indicative of negative sentiment?



# Machine Learning-Based vs. Lexicon-Based Sentiment Analyzer Accuracies (on CrowdFlower Test Set)



# Most Negative 10 Unigrams-

## General vs. Airline-Specific Tweet Training Data

(Bernoulli Naive Bayes Classifier)

General Tweet Training	Airline-Specific Training
1. Just	1. Cancelled
2. Miss	2. Service
3. Want	3. Hours
4. Like	4. Customer
5. Sorry	5. Hold
6. Time	6. Time
7. Love	7. Just
8. Know	8. Delayed
9. Really	9. Help
10. Today	10. Hour

["what", "questions", "do", "you", "have", "?"]