Airline Sentiment Analysis

Samantha Chu

Problem Statement

Airlines would like to understand how their consumers feel about their services and experiences

Given a set of reviews from Twitter, airlines would like to see how well they are serving their customers

Predict the sentiment of airline customers based on their twitter review

01

Data Collection

Data was collected from Kaggle. ~ Twitter data was scraped from February of 2015, for 5 major U.S. Airlines

02

Cleaning

Text column

Numerical columns

03

Modeling

Random Forest

AdaBoosting

04

Evaluation

Accuracy scores

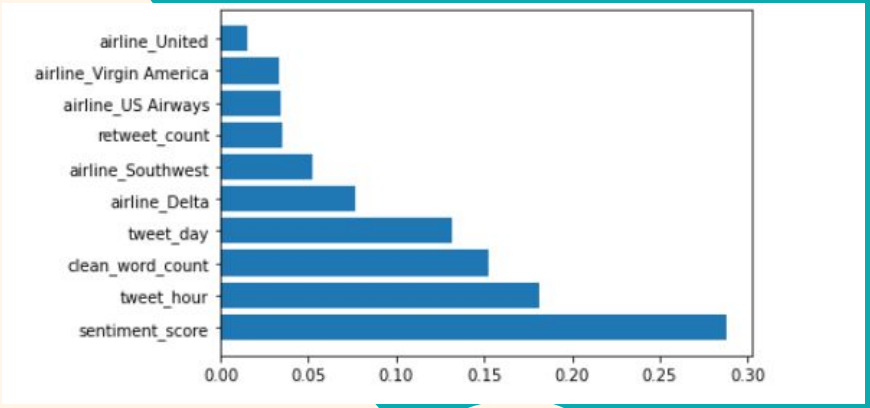
Feature Importances

negative	0.710114
neutral	0.165570
positive	0.124316

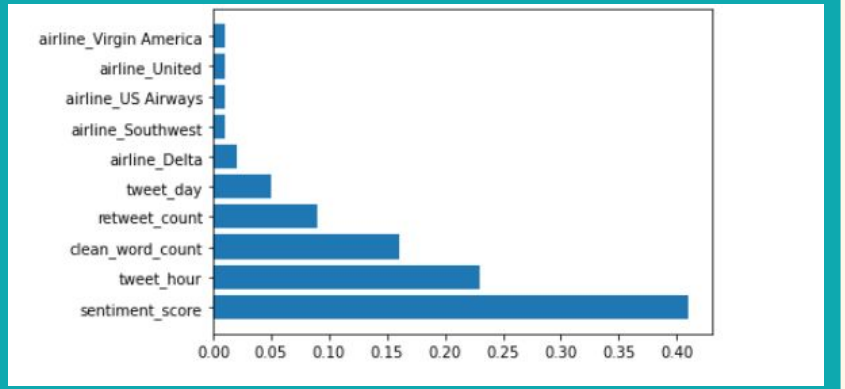
Results - Numerical data

	Model	Train_acc	Test_acc
0	RnadomForest	0.714833	0.713756
1	AdaBoost	0.783189	0.776047

Random Forest



AdaBoost

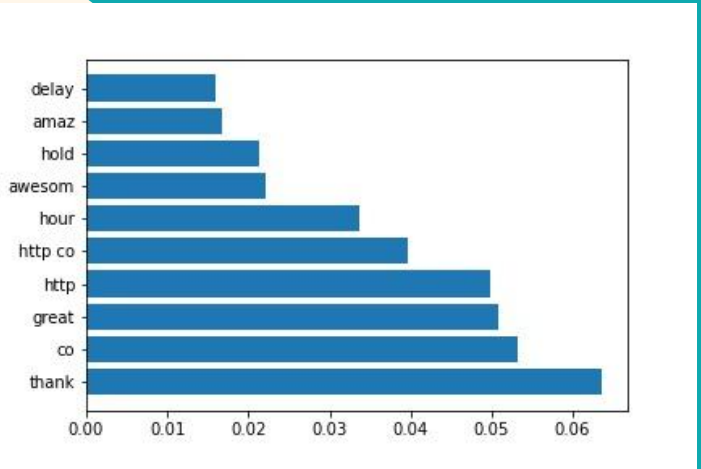


negative	0.710114
neutral	0.165570
positive	0.124316

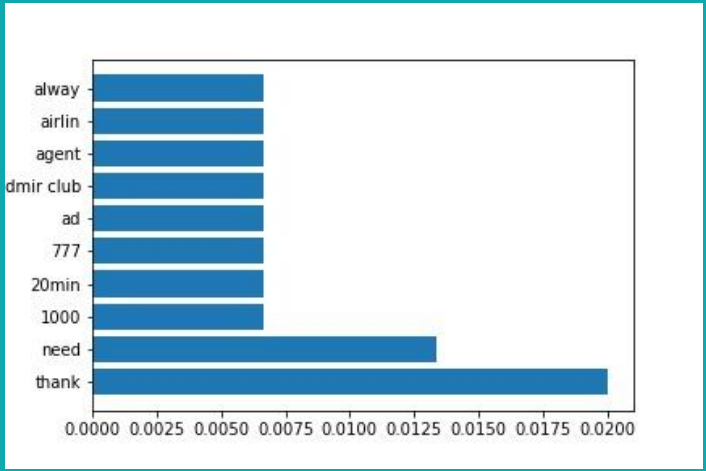
Results - Text data

	Model	Train_acc	Test_acc
0	RnadomForest	0.762794	0.720430
1	Adaboost	0.726452	0.717093

Random Forest



AdaBoost



Conclusions

- Use AdaBoost Model on text data
- Words such as: need, thank, great, awesome, delay were very predictive
- Basic Sentiment analysis is pretty powerful

Sources

Kaggle

- https://www.kaggle.com/crowdflower/twitter-airline-sentiment