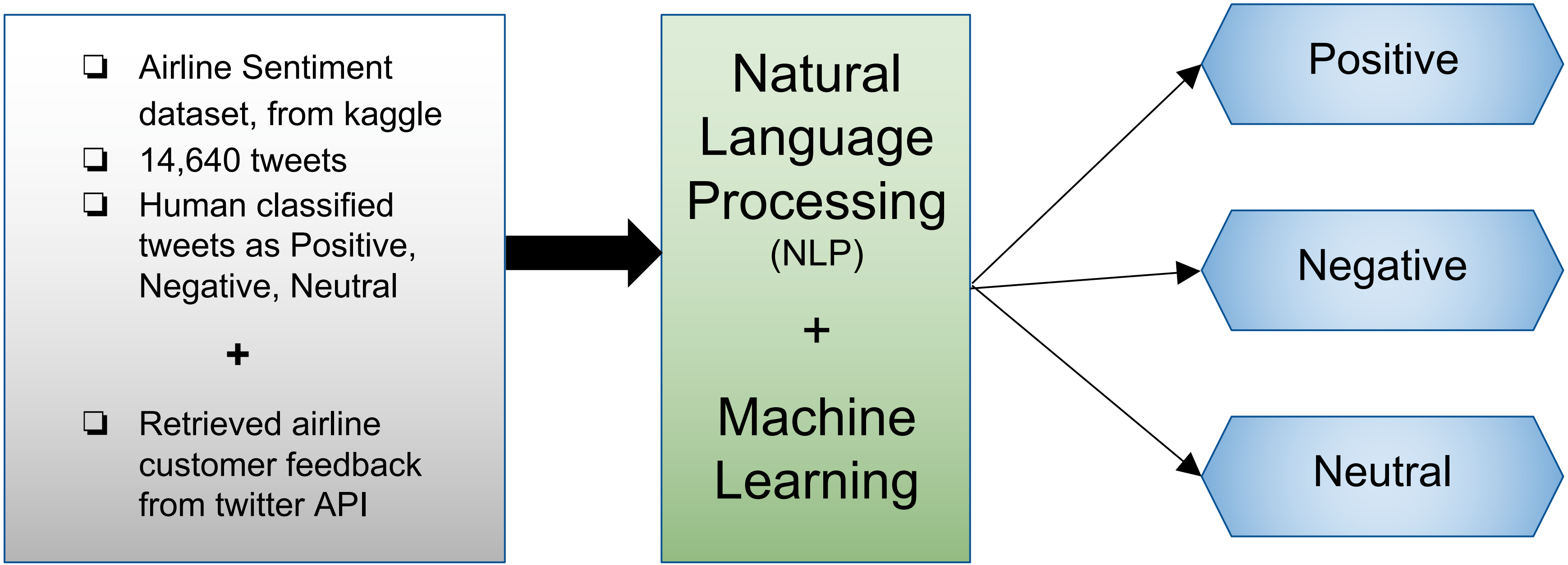

Airline Tweets
Sentiment Analysis
American
Alamy

Can sentiment analysis and machine learning help airline companies gauge customer satisfaction?

😊😞

By Sangita Gupta

Airline Sentiment Classification



Conversion of tweets into airline sentiment

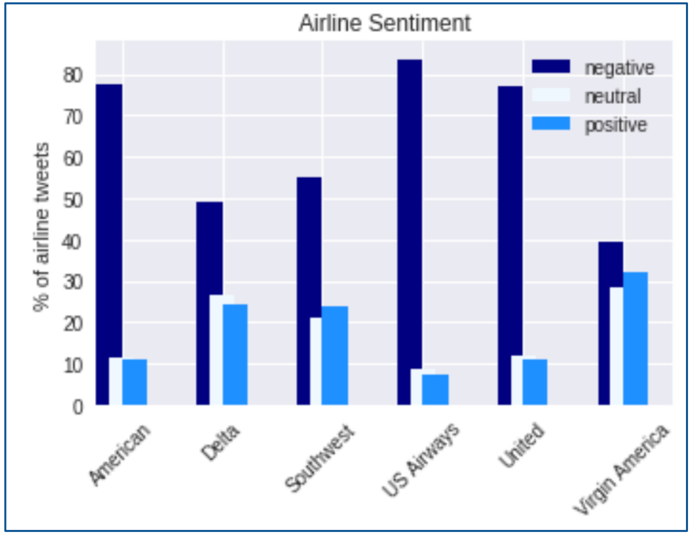
I ❤️ flying @VirginAmerica. ☺👍

👍👍✈️✈️💗 When are you guys going to start flying to Paris? @VirginAmerica: @LizaUtter You're welcome."

@VirginAmerica awaiting my return phone call, just would prefer to use your online self-service option :(

@VirginAmerica plz help me win my bid upgrade for my flight 2/27 LAX--->SEA!!! 🍷👍💺✈️

NLP
+
Machine Learning



Natural Language Processing

- Natural language processing includes many different techniques for interpreting human language.
- NLP methods used ⇒ tokenization, stop words removal, lemmatization and stemming.

Processing Tweets:

Raw Tweets

Clean Tweets: remove urls, @users, numbers, punctuation, …

Remove Stop Words: I, you, some, so, to, …

Apply Lemmatization: (find root of word)		Apply Stemming: (trim ends of word)
am, are, is ⇒ be better, great, best ⇒ good	OR	caresses, ponies, cats ⇒ caress, poni, cat

Process Tweets

- **Raw Tweet:**
  - @VirginAmerica Applied for Status Match on Feb 1. Got confirmation email same day. Still no news though. You guys have dropped ball Late Flightly 😢
- **Cleaned and Emoji Encoded Tweet:**
  - applied for status match on feb got confirmation email same day still no news though you guys have dropped ball late flightly emoji_34
- **Stop Words Removed Tweet:**
  - applied status match feb got confirmation email day news guys dropped ball late flightly emoji_34
- **Stemming Applied to Tweet:**
  - appli statu match feb got confirm email day news guy drop ball late flightli emoji_34
- **Lemmatization Applied to Tweet:**
  - apply status match feb get confirmation email day news guy drop ball late flightly emoji_34

Explore Emojis as sentiment predictors

Extract emojis from tweets to see how they relate to the classified sentiment.

emojis	❤☺👍	😡	😢	💜✈	🍷👍💺✈	💕💕	😁	❤	👏	😂💗	🍸	😒	👎	👍👍✈✈💗	😊😀😃😄	😎
airline_sentiment	positive	negative	negative	positive	neutral	positive	negative	positive	positive	positive	positive	negative	negative	neutral	neutral	positive

They look like good sentiment predictors. They are symbols so we have to encode them.

Users tend to group emojis together.

`@VirginAmerica plz help me win my bid upgrade for my flight 2/27 LAX--->SEA!!! 🍷👍💺✈️`

So first separate the emojis into individual symbols.

`@VirginAmerica plz help me win my bid upgrade for my flight 2/27 LAX--->SEA!!! 🍷 👍 💺 ✈️`

Now that they are separate, encode them as features.

Encode Emojis

❤️ : EMOJI_1	☺️ : EMOJI_2	👍 : EMOJI_3	😡 : EMOJI_4	😢 : EMOJI_5	💜 : EMOJI_6	✈️ : EMOJI_7	🍷 : EMOJI_8
💺 : EMOJI_9	😊 : EMOJI_10	😍 : EMOJI_11	👌: EMOJI_12	💕 : EMOJI_13	🌞 : EMOJI_14	😃 : EMOJI_15	😩 : EMOJI_16
😭 : EMOJI_17	😎 : EMOJI_18	🙈 : EMOJI_19	😁: EMOJI_20	❄️ : EMOJI_21	👏 : EMOJI_22	😂 : EMOJI_23	💗 : EMOJI_24
🍸 : EMOJI_25	😒 : EMOJI_26	👎 : EMOJI_27	😀: EMOJI_28	😄 : EMOJI_29	😘 : EMOJI_30	👸 : EMOJI_31	🎀 : EMOJI_32

- ❑ Encode emoji symbols as text.
- ❑ It will then get transformed to numerical data, along with the text data from the tweet.
- ❑ The text data needs to be transformed into numerical data, in order to train machine learning models.

Transform text into numerical data

Divide tweets into words and create a document term matrix. The columns are unique words in the corpus of tweets. The rows are tweets represented by a sequence of ‘0s’ and ‘1s’, depending on whether the word is present or not.

Tokenize Text

Corpus of tweets are separated into words

Corpus is now like a “bag of words”

Document Term Matrix

The “bag of words” is used to make a table

Columns are the unique words in the corpus

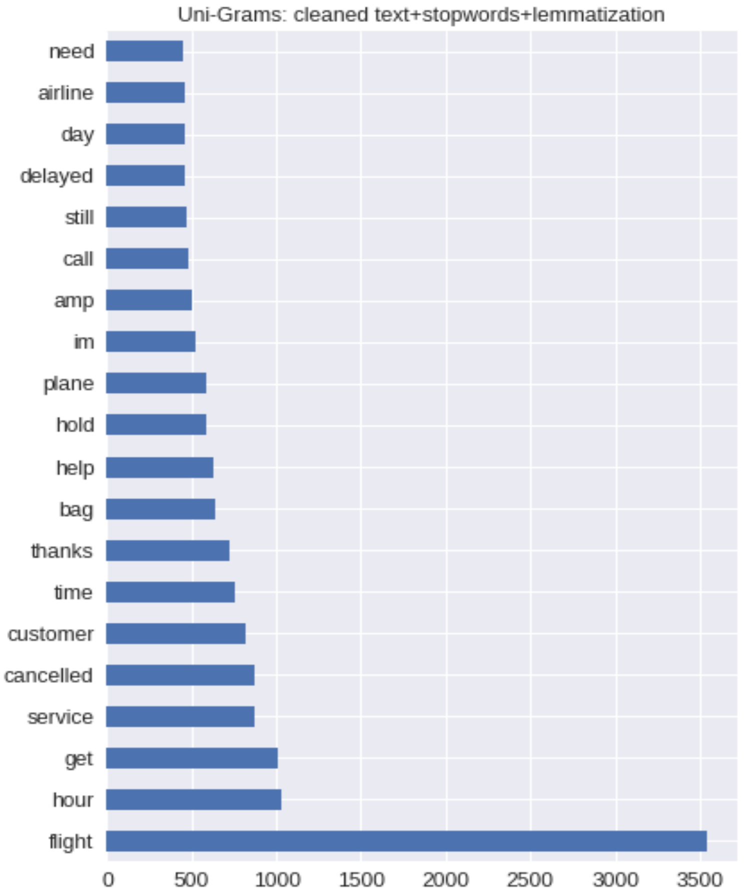
Rows are the individual tweets represented in numerical form

Vectorized Text

Each tweet is now represented by a series of ‘0s’ and ‘1s’ in the document term matrix

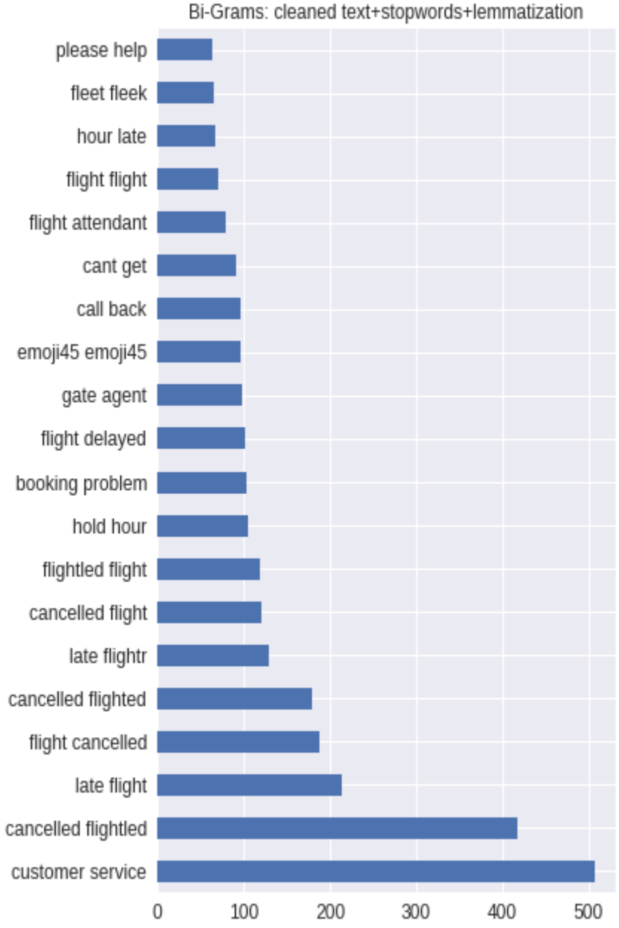
If the word exists in the tweet its value is ‘1’ if not, its ‘0’

Word Features: n-grams



- ❑ Given a sentence, we can construct a list of n-grams by finding groups of words that occur next to each other.
  - ❑ An n-gram is a contiguous sequence of n words.
  - ❑ There can be 1-grams, 2-grams, 3-grams…
  - ❑ These are also referred to as uni-gram, bi-gram, tri-gram.
- ❑ In Natural Language Processing, n-grams become features of the text data.
- ❑ N-grams are then used to perform text analytics, such as computing their occurrence frequencies throughout the corpus of text data.
- ❑ The graph to the left shows the top 20 “1-grams”, also known as “uni-grams”, for tweets in the dataset.
  - ❑ flight, hour, get, service cancelled, customer, time, thanks, bag, help, hold, plane, im, amp, call, still, delayed, day, airline, need

Word Features: bi-grams and tri-grams

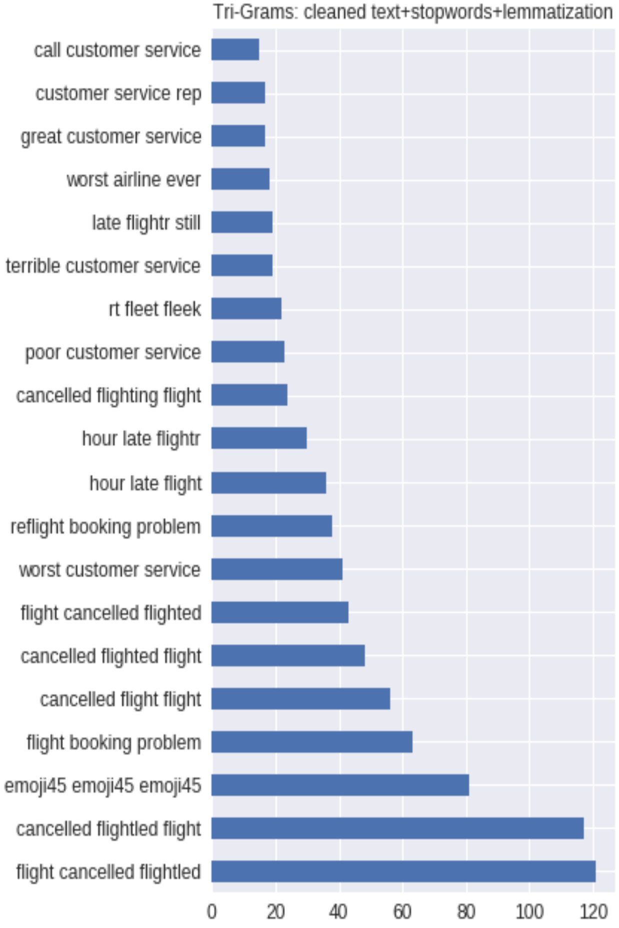


Top bi-grams:

- ❑ Customer service
- ❑ late flight
- ❑ flight cancelled
- ❑ hold hour
- ❑ booking problem
- ❑ flight delayed
- ❑ gate agent
- ❑ emoji45 emoji45
- ❑ hour late
- ❑ please help

Top tri-grams:

- ❑ Flight cancelled flightled
- ❑ cancelled flightled flight
- ❑ emoji45 emoji45 emoji_45
- ❑ flight booking problems
- ❑ cancelled flight flight
- ❑ worst customer service
- ❑ great customer service
- ❑ reflight booking problem
- ❑ hour late flight
- ❑ call customer service



Word Frequency: for sentiment classified tweets

Positive Classification

Negative Classification

Model Selection

Naive Bayes

Learning Mechanism

- Called a generative model.
- Learns a model of the joint probability, p(x,y), of the inputs x and label y, from the training data.
- Makes predictions by using Bayes rules to calculate the posterior probability p(y|x), and then picks the most likely label y.

Model Assumptions

- Assumes that every word in a sentence is independent from the other words.
- So for Naive Bayes the following sentences would all be the same.

"this was a fun party"
"this party was fun"
"party fun was this"

Logistic Regression

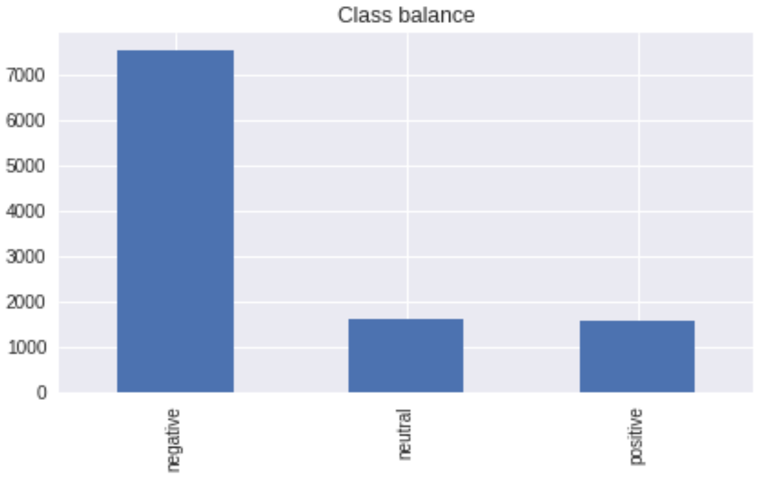
Learning Mechanism

- Called a discriminative model.
- Directly models the posterior probability distribution p(y|x), by learning the mapping from the inputs x to the labels y, from the training data.

Model Assumptions

- The independent variables x, are not linear combinations of each other.

Modeling



- Data is highly unbalanced.

Baseline Accuracy = 0.70

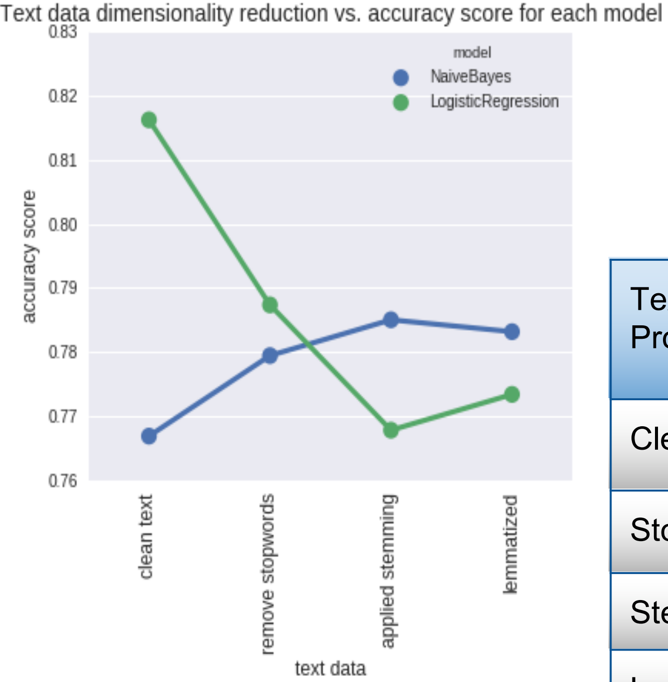
- We will be right 70% of the time if we always guess the most common class

Insights we want to gain

- ❑ What effects do ***emojis*** have on model predictions?
- ❑ What effects do different ***dimensionality reductions methods*** have on model predictions?
  - ❑ Stop Words Removal, Lemmatization/Stemming
- ❑ What effects does ***tuning*** the model's hyper parameters have on model predictions?

Model Accuracy Effects: using emojis as sentiment predictors

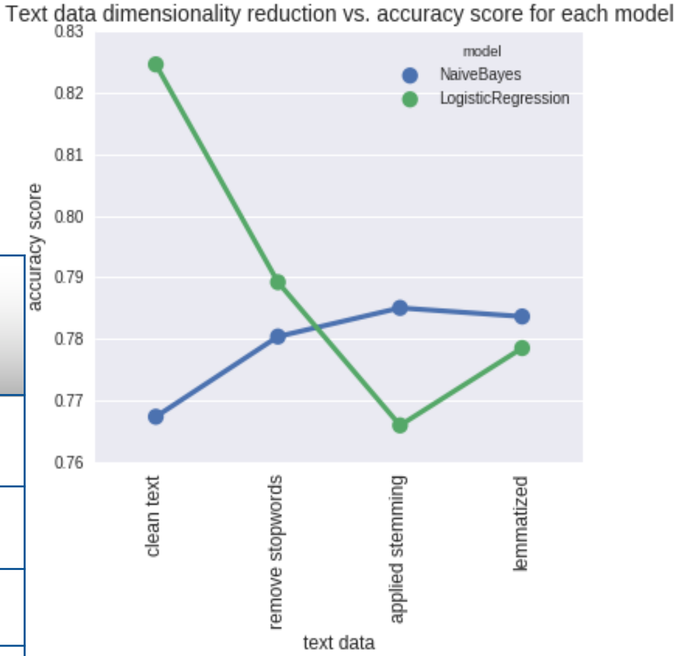
Emoji Not Encoded



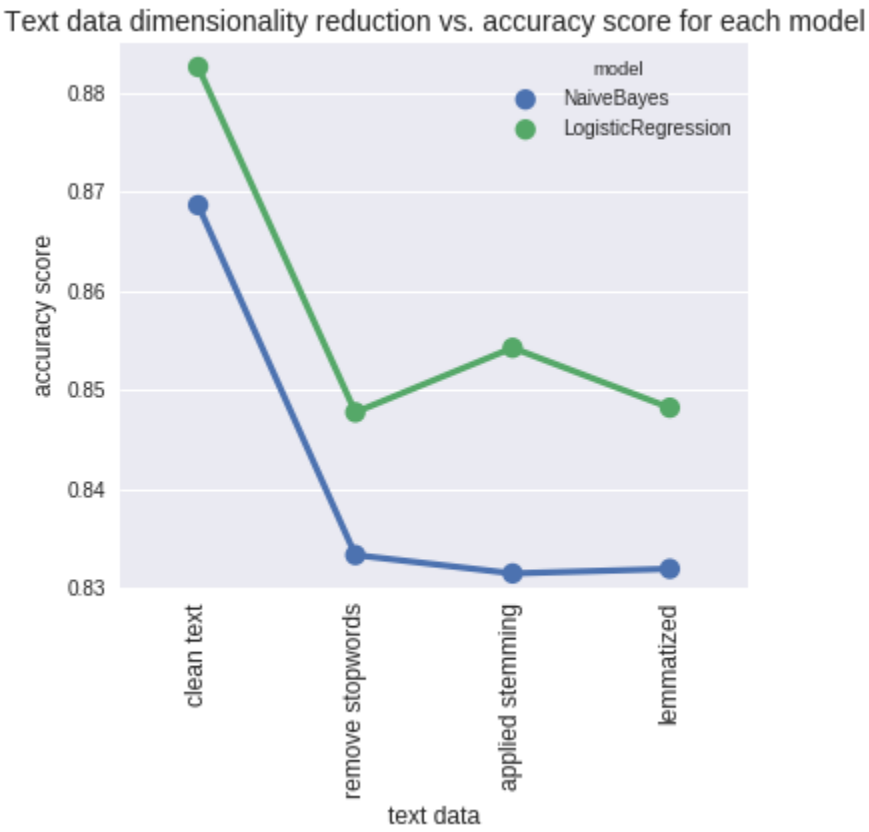
Model accuracy effects from encoding emojis

Text Processing	Naive Bayes	Logistic Regression
Clean Text	+ 0.047%	+ 0.836%
Stop Words	+ 0.093%	+ 0.186%
Stemming	+ 0%	- 0.186%
Lemmatization	+ 0.046%	+ 0.51%

Emoji Encoded



Model Accuracy Effects: tuning model hyper-parameters



Parameters Tuned:

- Countvectorizer
  - min_df = 5, max_df = 0.95, ngram_range = (1,2)
- Logistic Regression model
  - C = 1.0
- Naive Bayes model
  - Alpha = 0 .1

Model Accuracy effects from tuning models

Text Processing	Naive Bayes	Logistic Regression
Clean Text	+ 10.12%	+ 5.803%
Stop Words	+ 5.292%	+ 5.850%
Stemming	+ 4.642%	+ 8.82%
Lemmatization	+ 4.828%	+ 6.9642%

Best tuned model results

Model Accuracy of tuned models for different NLP methods, which reduce data dimensionality

Text Processing (Emoji Encoded)	Naive Bayes	Logistic Regression
Clean text	86.7%	88.3%
Stop Words	83.3%	84.8%
Stemming	83.1%	85.4%
Lemmatization	83.1%	84.8%

No data dimensionality reduction:

- ❑ Logistic Regression achieves an accuracy of 88.3%.
- ❑ Naïve Bayes achieves an accuracy of 86.7%.

With data dimensionality reduction:

- ❑ Logistic Regression achieves an accuracy of about 85%.
- ❑ Naïve Bayes achieves an accuracy of about 83%.

Model performance compared to baseline accuracy:

- ❑ Baseline accuracy was 70%, if always guessing the most common class, which in this case is the negative class.
- ❑ Logistic Regression achieves 18% better accuracy without dimensionality reduction and 15% better with.
- ❑ Naive Bayes achieves 16.7% better accuracy without dimensionality reduction and 13% better with.
- ❑ The machine learning models outperform the baseline accuracy by 13 to 18 percent.

Conclusions about Models

- **Using encoded emojis as sentiment predictors:**
  - Logistic Regression model accuracy seems to improve.
    - However, when using stemming accuracy actually decreases.
  - Naive Bayes model accuracy seems to marginally show improvement. For the most part it seem to be unaffected.
- **Tuning model hyper-parameters:**
  - Logistic Regression and Naive Bayes have significant accuracy gains by tuning hyper-parameters.
  - Naive Bayes has the highest accuracy gain when using data that is only cleaned and emoji encoded.
  - Logistic Regression has the highest accuracy gain when using data with stemming applied.
- **Applying data dimensionality reduction methods:**
  - Logistic Regression model accuracy seems to be about 3% worse than without dimensionality reduction.
  - Naïve Bayes model accuracy seems to be about 4% worse than without dimensionality reduction.

Classify airline tweets from Twitter

Airline customer tweets

TWITTER API

Pull tweets for major US airlines using hashtags

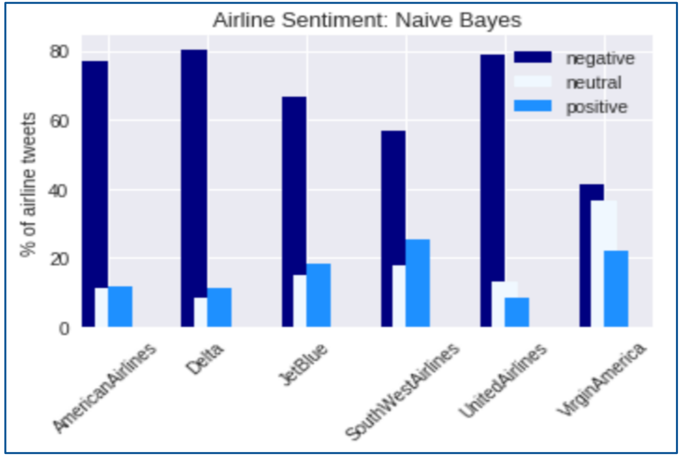
Process tweets

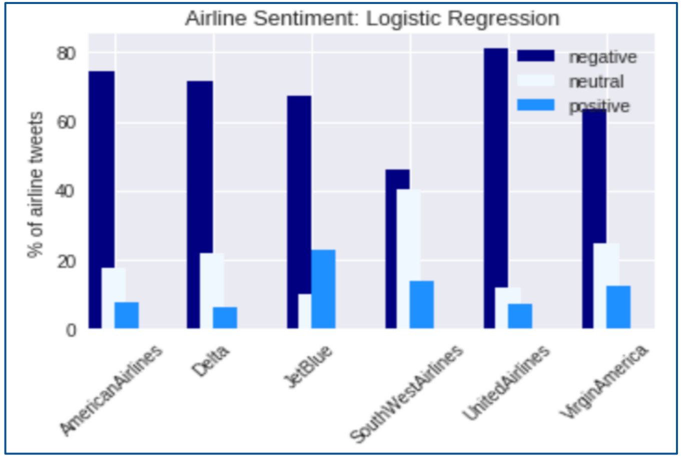
using

Natural Language Processing Techniques

Classify using Naive Bayes

Classify using Logistic Regression





Word Frequency: for sentiment classified tweets

Positive Classification

Negative Classification

Airline Sentiment Classification

Tweet: Kudos to the #unitedairlines staff for helping us with our crazy reservations. Nice send-off from CHS
Sentiment Prediction NB: positive **Sentiment Prediction LR:** positive

Tweet: @united YESSSSSSSSSSSSSSSSSSSSSSSSS! I don't know how today can get any better, this is amazing! #ThankYou #UnitedAirlines #FlyingTheTahitiSkies
Sentiment Prediction NB: positive **Sentiment Prediction LR:** positive

Tweet: A peek inside Classified, where CEOs and celebrities dine in a hidden restaurant at Newark Liberty International Airport. https://t.co/bpmhavKTSh #TableReady #speakeasy #UnitedAirlines #exclusive
Sentiment Prediction NB: neutral **Sentiment Prediction LR:** neutral

Tweet: @united That's too bad. 😞 #YYJ would love some more #UnitedAirlines service choices✈️
Sentiment Prediction NB: positive **Sentiment Prediction LR:** negative

Tweet: Montreal, Canada to Phoenix, Arizona for only $271 CAD roundtrip with United. #UnitedAirlines #Montreal https://t.co/3PErNhmXZ0
Sentiment Prediction NB: negative **Sentiment Prediction LR:** neutral

Tweet: I am really pissed off with @united Luggage missing since Jan 3 and no one from the airline has made any attempt to explain the problem #unitedairlines does not care
Sentiment Prediction NB: negative **Sentiment Prediction LR:** negative

Conclusions

- Sentiment analysis can be extremely useful in social media monitoring, as it allows us to gain an overview of the wider public opinion behind certain topics or brands.
- However sentiment analysis is not a perfect science.
- The human language is complex and teaching a machine to analyze the various grammatical nuances, cultural variations, slang and misspellings that occur in social media text, is a difficult process.

Challenges in predicting sentiment from tweets

- Machines learn best with precise, unambiguous and structured data. Tweets however, are not generally precise, often ambiguous and the language used is highly unstructured.
- Teaching a machine to understand how context can affect tone is even more difficult.
  - Humans are fairly intuitive when it comes to interpreting the tone of a piece of writing.
    - Consider the following sentence: “My flight’s been delayed. Brilliant!”
    - We can easily identify the sentiment as negative, however without contextual understanding a machine looking at the sentence above might see the word “brilliant” and categorize it as positive.

Next Steps

1 - Dealing with Class imbalance in the dataset

To improve model prediction accuracy, I will explore techniques of dealing with class imbalance.

- ❑ The dataset is highly imbalanced in favor of the negative class.
- ❑ Resample the most frequent class to have a similar corpus size as the other classes.
- ❑ Tune the penalty hyper-parameter of Logistic Regression.
- ❑ Evaluate other algorithms which deal well with imbalanced datasets.

2 – Explore topic modeling and content categorization techniques

- ❑ Infer topics from sentiment classified tweets to gain further insights on public opinion.

Questions?

To test out my airline sentiment classifiers go to:

http://34.212.204.117:5000/predict-sentiment-interface