

Análisis de Sentimientos a través de Tweets de Aerolineas de E.U.A.

ARIEL ELIEZER MIGUEL VILLEDA

Contenido

- Introducción del proyecto
- Manejo de los Datos
- Demostración de primeras pruebas

Explicación del Proyecto

Introducción

- El análisis de sentimientos (sentiment analysis) es el estudio por el cual se determina la opinión de las personas en Internet sobre algún tema en específico
- Twitter se ha convertido en una de las plataformas en línea más utilizadas para expresar opiniones e ideas

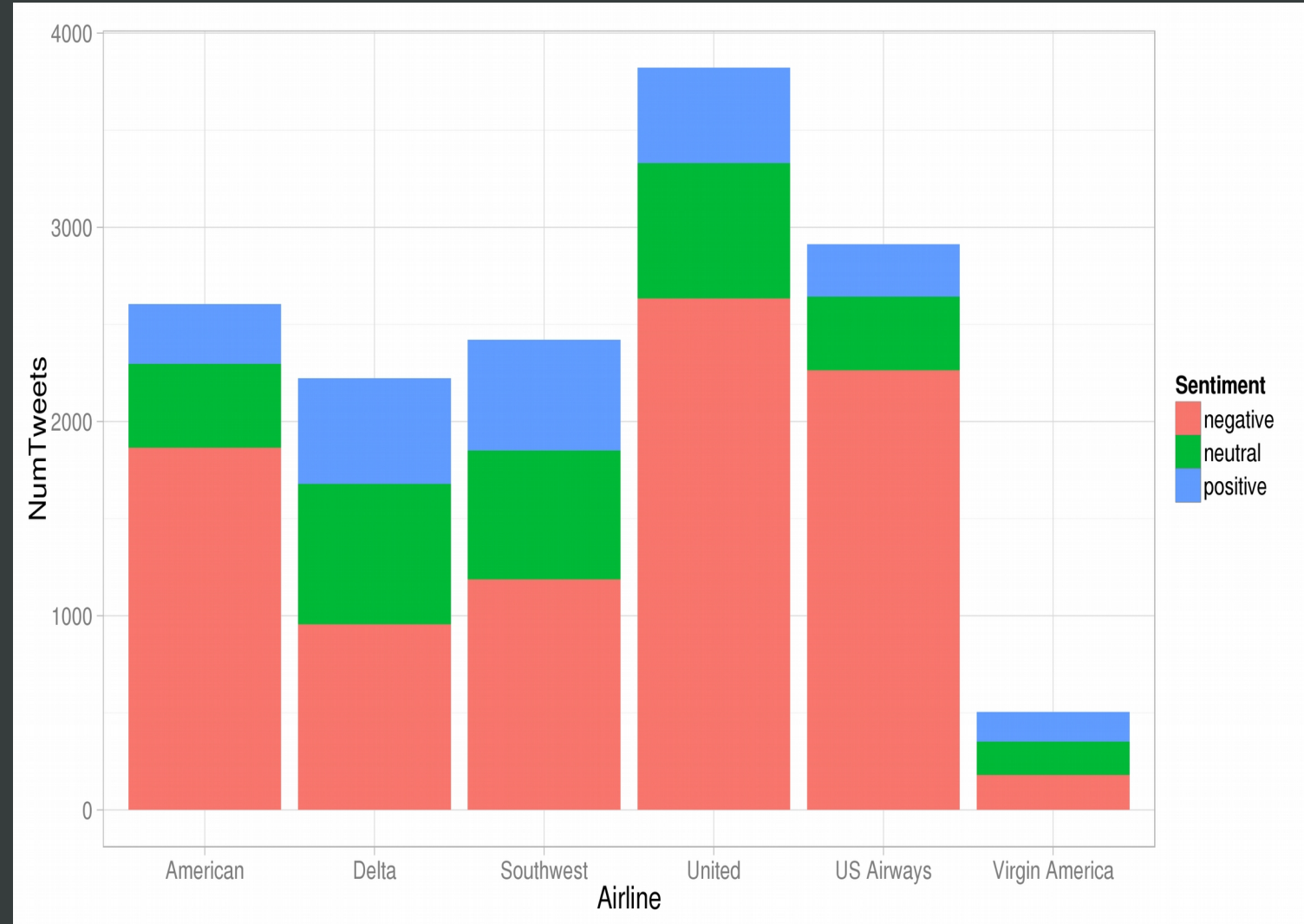
Objetivo

- Analizar sentimientos acerca de los problemas en las principales aerolíneas de Estados Unidos
- Generar un modelo basado en análisis de n-gramas y aprendizaje automático
- Entrenar el modelo con un conjunto de datos ya categorizado del año 2015

Manejo de los Datos

Características de los datos

- Los datos de entrenamiento se encuentran categorizados en negativos, neutrales y positivos
- Si se clasifica “negativo”, el vector contiene una categorización de “causa”, como puede ser “vuelo retrasado”, “mal servicio”, etc. de valores normalizados



Características de los datos

- Un ejemplo de un vector (solo características de posible interes):

airline_sentiment	airline	retweet_count	text	tweet_created
positive	Virgin America	0	@VirginAmerica plus you've added commercials to the experience... tacky.	2015-02-24 11:35:52 -0800

Demostración de pruebas

Algoritmos Utilizados

- Librería python nltk:
 - Repertorio de 'stopwords' para el idioma Inglés
 - Función TweetTokenizer()
- Librería python sklearn:
 - SGDClassifier para el modelo de clasificación de sentimiento en los Tweets
 - CountVectorizer clase para el conteo de conteo de tokens el tweet
 - Método fit_transform() que genera una matriz “término-documento” aprendiendo el vocabulario

¿Qué sigue?

- Agregar de manera Automática, la “confianza en el sentimiento” de acuerdo al corpus del Tweet para futuros entrenamientos del modelo
- De manera automática, dar seguimiento al “hilo” del Tweet (ver respuestas al mismo) en busca de una tokenización más efectiva
- Experimentar agregando otras combinaciones de características al vector

Gracias por su atención