US Airlines Twitter Sentiment Analysis

Dasun Wellawalage
DSC 680
https://github.com/dasun27/Data_Projects

Domain

- https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/

  This article covers the sentiment analysis of any topic by parsing the tweets fetched from Twitter using Python.

- https://towardsdatascience.com/twitter-sentiment-analysis-classification-using-nltk-python-fa912578614c

  This is a Twitter sentiment analysis attempt using natural language processing techniques.

- https://www.pluralsight.com/guides/building-a-twitter-sentiment-analysis-in-python

  Another Twitter sentiment analysis using Python based machine learning libraries.

- https://medium.com/@francesca_lim/twitter-u-s-airline-sentiment-analysis-using-keras-and-rnns-1956f42294ef

  This is a Twitter U.S. Airline Sentiment Analysis using Keras and RNNs.

- https://www.datasciencecentral.com/profiles/blogs/sentiment-analysis-of-airline-tweets

  This is a detailed guide on sentiment analysis for airline tweets.

- https://ipullrank.com/step-step-twitter-sentiment-analysis-visualizing-united-airlines-pr-crisis/

  This post goes into great lengths to explain not only the sentiment analysis process but also how to create an application to collect user sentiments.

- https://www.kaggle.com/parthsharma5795/comprehensive-twitter-airline-sentiment-analysis

  This is a notebook analyzing the same dataset found on Kaggle. He uses a classification approach for his analysis.

- https://www.kaggle.com/anjanatiha/sentiment-analysis-with-lstm-cnn

  This notebook tries to perform sentiment analysis using LSTM & CNN.

- https://www.kaggle.com/mrisdal/exploring-audience-text-length

  Another effort to use audience & tweet length to perform sentiment analysis.

- https://www.kaggle.com/langkilde/linear-svm-classification-of-sentiment-in-tweets

  This notebook uses Linear SVM classification for sentiment analysis.

Data

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

There is no codebook. Below are the variables.

- tweet_id
- airline_sentiment
- airline_sentiment_confidence
- negativereason
- negativereason_confidence
- airline
- airline_sentiment_gold
- name
- negativereason_gold
- retweet_count
- text
- tweet_coord
- tweet_created
- tweet_location
- user_timezone

Research Questions? Benefits? Why analyze these data?

User feedback is quite important in any business and specially, in online businesses. Simply because other users read these feedbacks before deciding to use your product or service. Social media platforms like Twitter play a major role in advertising today. Thus, it is quite important to make sure the overall sentiment for your products and services remain positive on these platforms. By analyzing these sentiment data, we are trying to figure out the major reasons for a user to develop a negative impression on your products or services. Particularly, in this case we are studying the sentiments for US airlines so we can understand the positives and negatives in the US airline industry today.

Methods

Based on the dataset there are a few methods I can use to analyze this. There is a specific column that has already identified each sentiment as positive, negative, or neutral and there is confidence rating associated too. I can base my analysis on this and perform standard algorithms-based testing. Also, I can analyze the content in these tweets and try some natural language processing too. However, I am not sure if that will be feasible, given the time and resource constraints I have.

Potential Issues

This dataset is somewhat small (14k+) compared to most datasets. As a result, it might produce results that are not really useful in a real-world scenario. There are some missing values which I will have to manage during the analysis. I might consider combining this with another dataset but that depends on the

time and resources available. Apart from that, I don’t expect any other issues with this dataset.

Concluding Remarks

In the recent past we have seen a lot of negative sentiment towards the US airline service due to some controversial incidents. Not just in US, but overall, in the entire world, most people are dissatisfied with the airline service. Analyzing this dataset and identifying the customer pain points can help any airline gain a competitive advantage. Although the dataset is somewhat smaller for a comprehensive analysis, the basic approach and methods used in this analysis can be applied to a larger project too.