Sentiment Analysis: How is Your Brand Viewed on Twitter?

# Matt B. Jackson DSC 680 - Summer 2020 https://mbjackson.github.io/

# Domain/References

Bastani, K. (2019, September 19). *Sentiment Analysis on Twitter Data Using Neo4j and Google Cloud* . Retrieved from Kenny Bastani: https://www.kennybastani.com/2019/09/sentiment-analysis-on-twitter-data.html

Gupta, S. (2018, January 7). *Sentiment Analysis: Concept, Analysis and Applications*. Retrieved from Towards Data Science: https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17

Joshi, P. (2018, July 30). *Comprehensive Hands on Guide to Twitter Sentiment Analysis with dataset and code*. Retrieved from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2018/07/hands-on-sentiment-analysis-dataset-python/

Maynard, D., & Greenwood, M. (2014). Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis. *The LREC 2014 Proceedings*, 26-31.

Nagpal, A. (2018, July 31). *Sentiment Analysis of Twitter using Spark*. Retrieved from CloudSigma: https://www.cloudsigma.com/sentiment-analysis-of-twitter-using-spark/

Nodarakis, N., Sioutas, S., Tsakalidis, A., & Tzimas, G. (2016). Large Scale Sentiment Analysis on Twitter with Spark. *EDBT/ICDT*, 1-8.

Pascual, F. (2019, June 7). *Twitter Sentiment Analysis with Machine Learning* . Retrieved from MonkeyLearn: https://monkeylearn.com/blog/sentiment-analysis-of-twitter/

RapidAPI. (2020, March 31). *How To Build a Sentiment Analysis App*. Retrieved from RapidAPI: https://rapidapi.com/blog/build-sentiment-analysis-app/

The Comprehensive R Archive Network. (2019, April 2). *Sentiment Analysis of Twitter Data*. Retrieved from The Comprehensive R Archive Network: https://cran.r-project.org/web/packages/saotd/vignettes/saotd.html

Zhou, X., Tao, X., Yong, J., & Yang, Z. (2013). Sentiment analysis on tweets for social events. *IEEE 2013*, 557-562.

**Data Source**

1. “Sentiment140 Dataset with 1.6 million tweets” – Kaggle
   1. Dataset consists of 1 CSV files each with 6 variables. The entire dataset takes up around 200 MB of data.
   2. https://www.kaggle.com/kazanova/sentiment140

# Research Questions

1. How can a model be created in Apache Spark to create a Sentiment Analysis model?
2. Can a model be integrated with Django in Python to create a real-time dashboard to view user sentiment on Social Media?

# Methods

In this project the dataset will be processed in Apache Spark by using a pipeline to chain the various functions to process text. From here the features will be used to train the model and test it on new data. Once the model is created it will be deployed by using Django in Python to create a real-time dashboard where brands can view how people feel about their brand in real-time. In order to do this the Twitter API will be used to stream in tweets.

# Potential Issues

The biggest issue I think I would have with this project is Spark. While I know it’s basic functions, I don’t feel as confident in using it to create a meaningful program. Because of this is why I wanted to use it to try and expand my skill set and have a good Spark project in my portfolio.

# Concluding Remarks

With the advent of Social Media, all brands need to be aware how they are being portrayed online. It been common these days to see a brand make a poor choice or ignore customer feedback which results in the deterioration of the brand. The end goal of this project is to create an easy way to view the sentiment associated with a brand in order to make the right business decisions.