

Capstone project 1

Sentiment analysis for products launched using Twitter data

The aim is to use Twitter data to perform sentiment analysis on products that are being launched globally.

- Tweepy to be used to scrap unstructured data from Twitter.
- Features of focus are time of tweet, tweet content, country of origin and number of retweets.
- As part of data wrangling we use
 - demoji library that helps in removing emoji in the tweets.
 - re (regular expression) library to remove unwanted chars like #,@, handle names etc.
 - Google Places API in order to identify countries, since location data usually contains partial data like just the name of city, state or the value is in native language. Missing values will be marked as unknown.
- With the data cleaned, Textblob library is used for sentiment analysis.
- Other visualizations can be performed using either matplotlib, Seaborn or wordcloud draw up basic outline of the data.
- Using the stream method we capture tweets over a particular duration to and count the tweets for each product to get a basic idea of which one is more popular.

Note:

- Twitter free developer account has limits for the number tweets collected over an interval.
- Sign up to Google APIs premium account (free for 1 year) will be required. Although this contains a monthly usage limit.