Twitter Scraping for Sentiment Analysis

Client: A popular electronics manufacturer from Japan

Offering: Site-specific Crawling

Challenge:

The client was planning to do sentiment analysis on top of tweets mentioning their product or brand name. To achieve this, tweets mentioning their product/brand names had to be extracted along with the twitter handle, number of likes, number of retweets, hashtags used and the URL of the tweet.
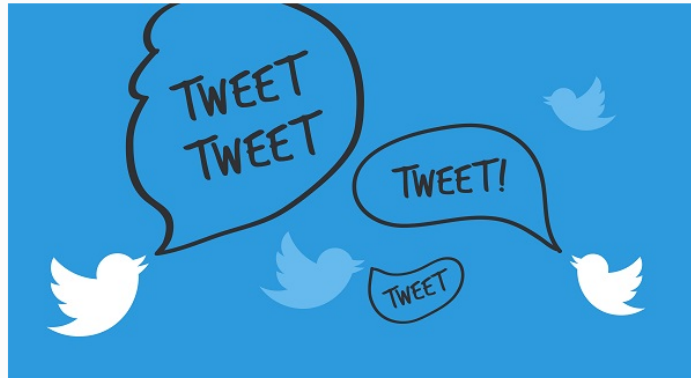
Solution:

Our team was provided with the list of keywords to be monitored while crawling twitter. The crawl had to be repeated every day to fetch new records. For this, our team programmed a crawling setup that could monitor twitter for the given set of keywords and fetch the posts that had these keywords. For every detected instance of the keywords, the crawler would extract the required data points. Although twitter has its own API, we had to use a custom crawler to effectively handle the requirement. The client opted for data to be delivered in XML format to their Dropbox account. It took only 2 days to complete the initial setup; after that the data flow started. The client could perform sentiment analysis using the provided datasets in a matter of few days.

Benefits to the client:

- All the technical aspects of data extraction were taken care of by our team

- The setup took only 2 days to complete and the data delivery was consistent thereafter

- We set up a monitoring system to spot website changes so as to promptly modify the crawler

- Our tech stack could handle huge amounts of data without any bottlenecks

- The client could improve their customer experience by using sentiment analysis and feedback

- The cost incurred was competitive and much lower than what an in-house crawling setup would have cost them