

Datasets

Here we have used the Twitter developer API to fetch a live stream of tweets.

To get the stream of tweets we have used a “Tweepy” stream listener to pool a stream of live tweets.

Tweepy Stream Listener: [http://docs.tweepy.org/en/v3.5.0/streaming_how_to.html#step-1-creating-a-streamlistener]

We are consuming real-time live tweets from the Twitter Stream API and asynchronously dump it into our KAFKA topic.