BIG DATA
&
INTELLIGENT ANALYTICS

Final Project

BY

Anirudha Deepak Bedre

Avikal Adeep Chhetri

(Team 3)

Twitter Sentiment Analysis

Problem Statement

To build a model that obtains and classifies the trend of sentiments of a stream of tweets for a given #hashtag

How are you feeling?

Why?

Public relations to track customer reviews

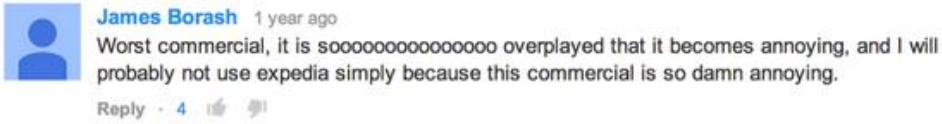
Politics

Events

Social media monitoring

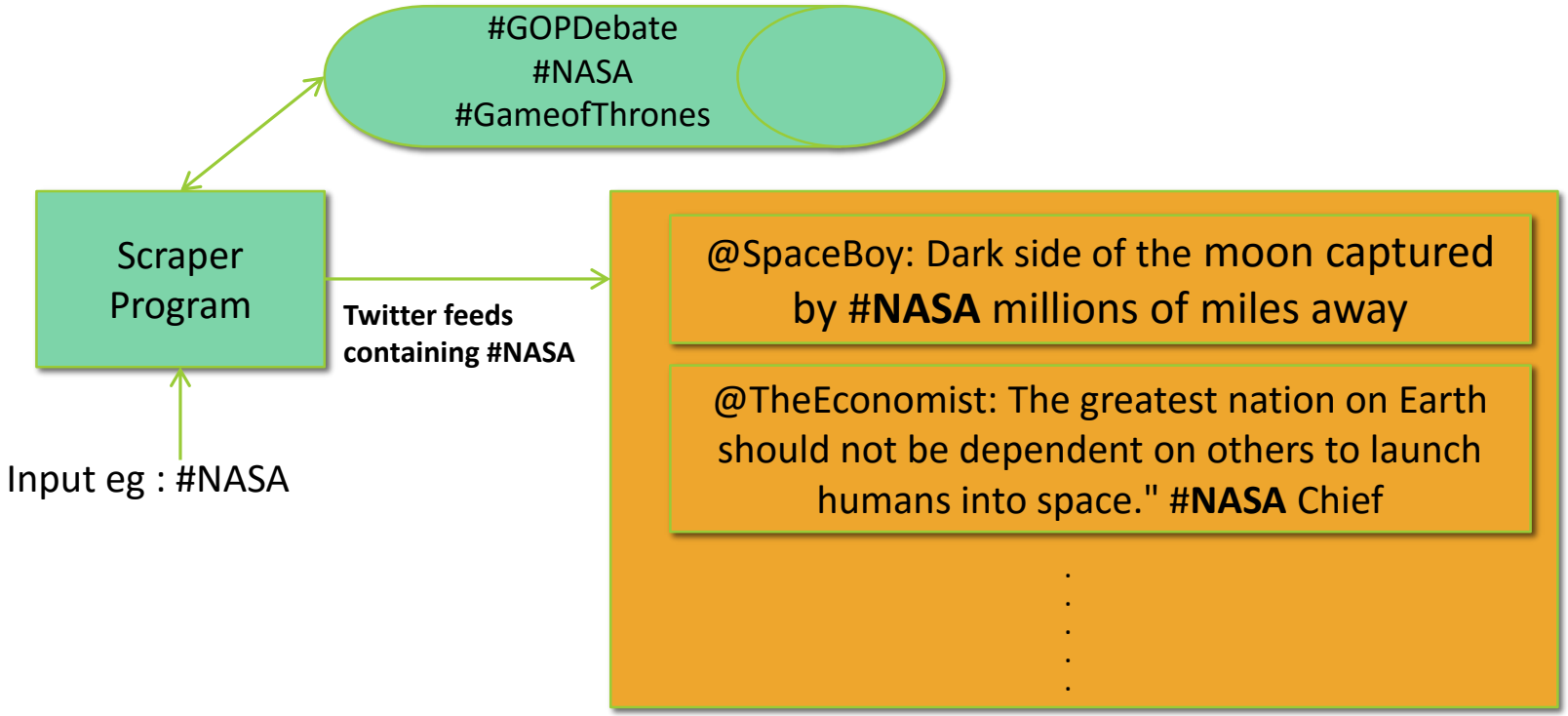
Customer Satisfaction Evaluation Metrics

News Channels



Dataset

Live Stream of Twitter feeds which are filtered by given #hashtags as input



Technologies And Software Used

Kafka 2.11-0.8.2.1

Twitter API 1.1

Apache Spark 1.4.0

IPython Notebook

Plotly 1.8.3

Approach

First, a Kafka Producer is used to fetch the tweets from the Twitter API by applying the #hashtag input as the filter. The producer then emits the tweets in Avro format to be received by the consumer.

We then have a Kafka consumer which receives the tweets emitted by the producer and process them by using the Spark Streaming context.

Now the stream of tweets are processed as ‘RDDs of tweets’. Each of the tweet text is now mapped and processed for sentiment analysis.

The sentiment analysis is calculated by accessing the Stanford Core NLP libraries and gives out the resultant sentiments like: ‘Negative’, ‘Very Negative’, ‘Neutral’, ‘Positive’, ‘Very Positive’.

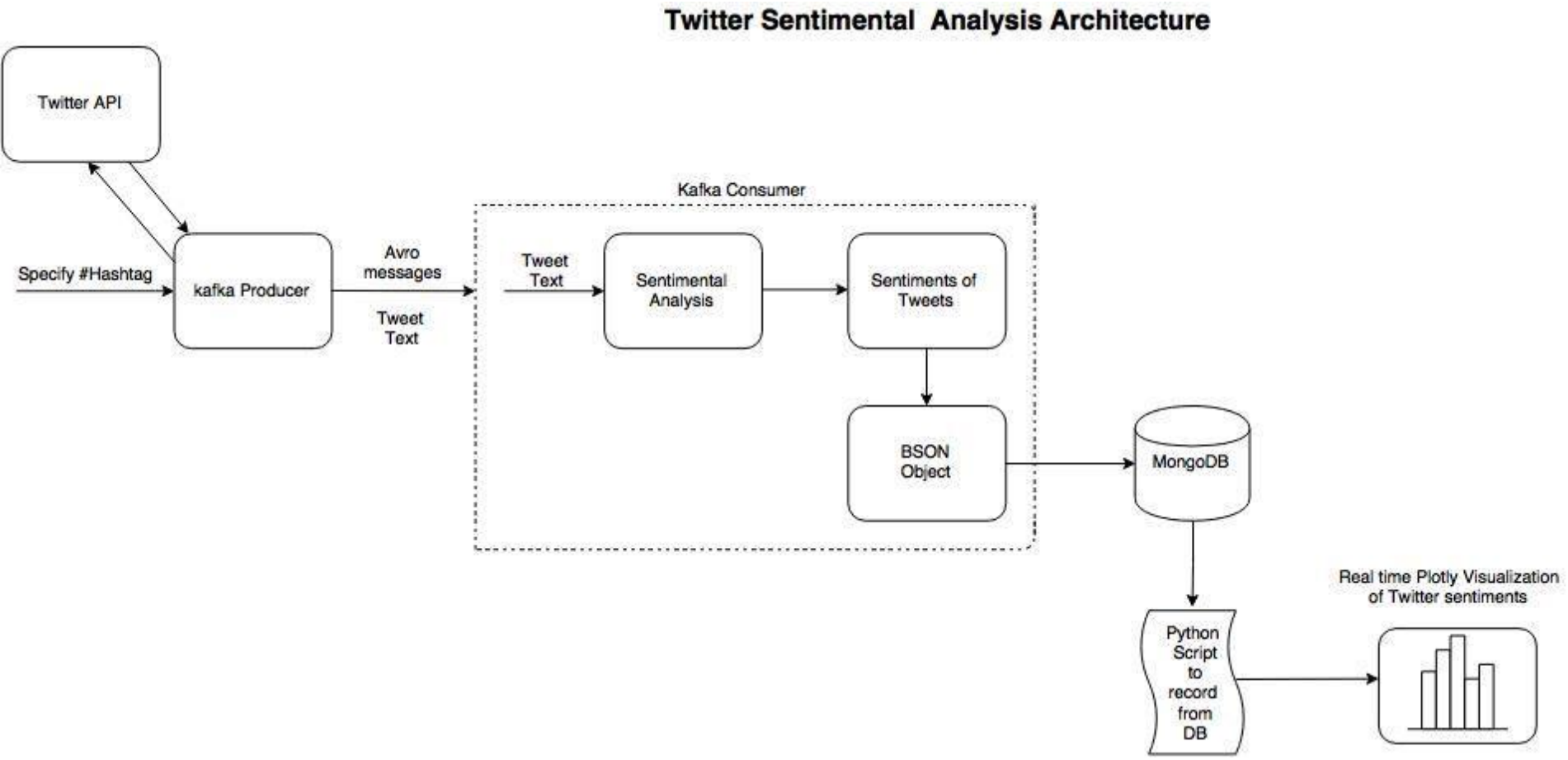
Approach (Continued)

The total sentiment count of each category is calculated and is stored in the database (MongoDB) by using BSON objects.

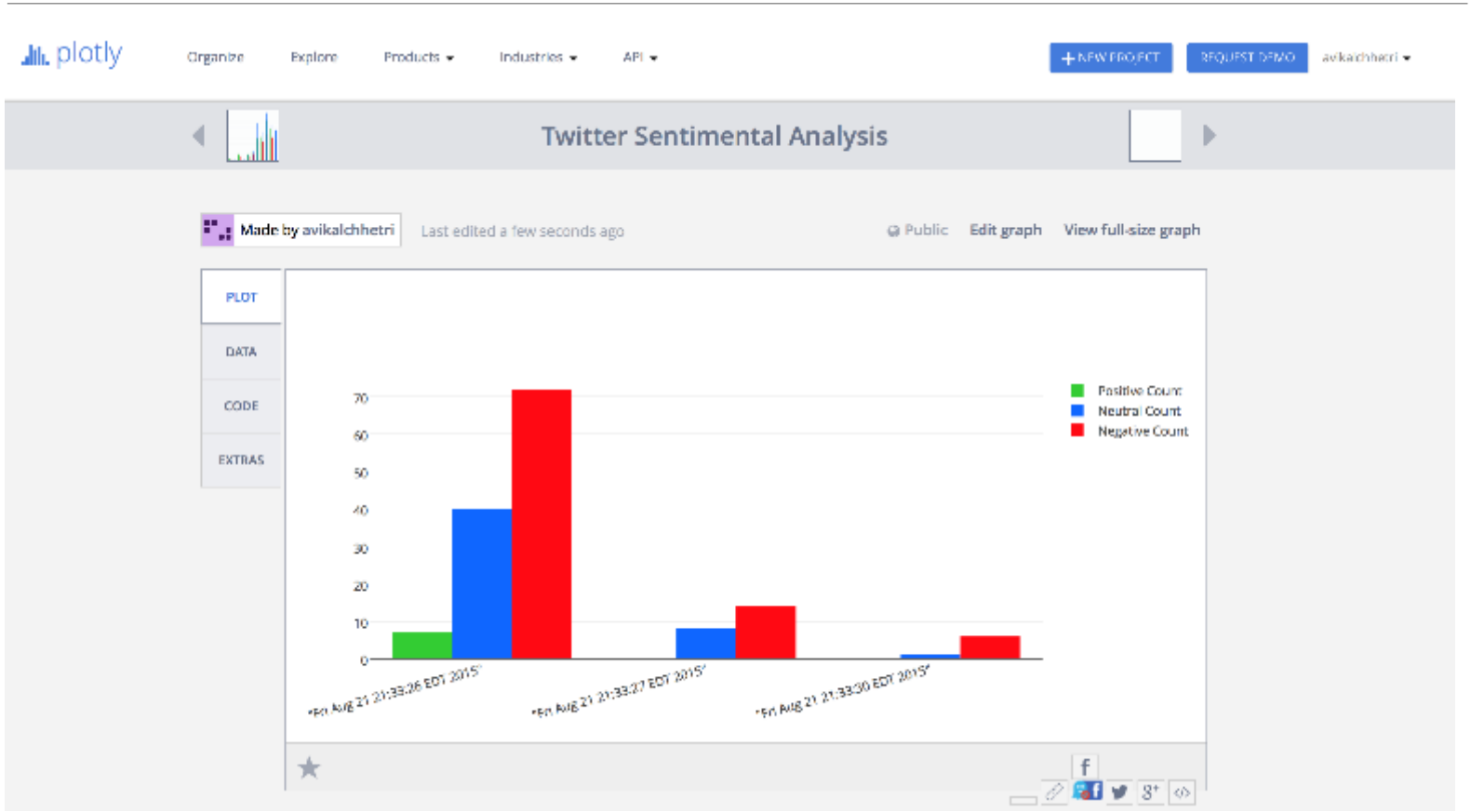
A python script is then run to access the records in the database in a timed loop and respectively stream the data to the Plotly servers.

Plotly, upon receiving the data displays a streaming graph of the sentiments observed.

Architecture Diagram



..... a sneak peek!


plotly
Organize
Explore
Products
Industries
API
+ NEW PROJECT
REQUEST DEMO
avikalchhetri
Twitter Sentimental Analysis
Made by avikalchhetri
Last edited a few seconds ago
Public
Edit graph
View full-size graph
PLOT
DATA
CODE
EXTRAS
70
60
50
40
30
20
10
0
Positive Count
Neutral Count
Negative Count

Demo

Thank you!