Team12

- Title: **Twitter Stream Categorizer**
- Description: The project aims at capturing and analyzing streaming data from Twitter to compare the different categories that come up in tweets. The various categories we want to take into consideration are political, social,entertainment, environmental, science & technology,finance and sports
- Team members:

Shriya Sharma	ssharm25
Sonal Patil	sspatil4
Sai Sameer Tirumalasetti	stiruma

Deliverables

1. Implement an end to end infrastructure pipeline that takes in twitter data stream into Kinesis, processes it and stores it into S3.
2. Perform analysis on the captured data independent of the volume of data being received from Twitter using Kinesis Analytics.
3. Giving out near real-time results for the analysis performed updating results of analysis with the changing input data by monitoring the traffic.
4. Define classifiers that categorize the data obtained from stream in application and build models based on them.
5. Generate graphical visualisation in the model to compare the different categories of the data captured.

Dependencies

The following are the dependencies that the mentioned project will be needing to carry out the desired tasks:

1. Amazon Kinesis Firehose
2. Amazon S3 (Storage)
3. H20/Amazon ML
4. Kibana
5. Dataset: Twitter stream
6. VCL

Issues

1. Would need a dedicated VCL instance.