#PrecogSummer 17 Tasks

Task A

Task A requires you to analyse the US Presidential Elections from a text point of view. We would like you to perform analysis on the tweets that you shall collect via the Streaming API by filtering the stream for #USelections. We expect you to curate at least 10,000 tweets and store them in a MongoDB collection.

From the tweets that you curate, we would like you to make a simple Web Application which showcases at least 4 of the following:

- Locations of the Tweet (Use any geoplotting library of your choice. A few to name are Google Charts, CartoDB)
- Who is more popular, Hillary or Trump? (You can create a metric for yourself, and define what it is in the applciaiton)
- List of Top 10 Hashtags being used in the stream
- Distribution of Original Tweets vs Retweeted Tweets
- Distribution of favorite counts on Original Tweets
- Distribution of Type of Tweet i.e. Text, Image, Text+Image

You would be required to upload your MongoDB collection into your Gitlab Repository. More details regarding the same will be conveyed to you near the end of the Programming Task Deadline.

Task B

Task B entails you to develop a Web Application, which when given an image, is able to give a result saying whether Donald Trump and/or Hillary Clinton are present in the image or not.

You are not allowed to use any online services and/or APIs for facial recognition, but are supposed to build your own model(s) from scratch. Feel free to use any image processing and learning techniques/libraries. You have the full liberty to decide on the architecture of the system.

You are required to curate a Training Set for your model(s) by using Twitter APIs (either REST or Streaming) and would be required to store them in a MongoDB collection. We expect you to use your curated Training Set to perform, at least, a 5 fold cross validation to judge for yourself how well your model(s) is/are working.

We will also be releasing our Test Set soon.

You would be required to upload your MongoDB collection into your Gitlab Repository. More details regarding the same will be conveyed to you near the end of the Programming Task Deadline.

Sample Input and Output have been explained on the next page.

Sample Input	**Sample Output**

	{
	“Donald Trump”: “No”,
	“Hillary Clinton”: “No”
	}

	{
	“Donald Trump”: “Yes”,
	“Hillary Clinton”: “No”
	}

	{
	“Donald Trump”: “No”,
	“Hillary Clinton”: “Yes”
	}

	{
	“Donald Trump”: “Yes”,
	“Hillary Clinton”: “Yes”
	}

Guidelines (for Web Deployment of Task A and Task B):

- Use a Python based Web Development Framework to deploy your Web Application. If you are starting with Python based Web Development Frameworks, check out Flask Framework. It's relatively easy to learn it's basics, and is a quite light framework.
- Deployment of Web Application has to be done on any one of the following platforms:
    - AWS EC2
    - Digital Ocean Droplet
    - Heroku App

  These platforms have free instances for students, should you face any difficulties in acquiring a free server from these platforms, do not hesitate to ping us.

A good tutorial on getting a simple flask application up and running can be found here.

Task C

Task C involves the development of either a Chrome Extension or an Android Application which gives the user a notification when any of the handles that have been added for tracking post a tweet. We expect you to develop a REST API which helps keep track of a user's handles to track by storing them in a database/text file and also integrate facebook login, so that if a user changes from platform to another, s/he shall be able to retrieve the handles added for tracking.

It would be wonderful if you can keep a copy of your progress in the form of an apk/crx file in your repository at all times.

Some twitter handles that we would like your Extension/Application to follow by default are @realDonaldTrump and @HillaryClinton.