Twitter Analysis Of ISIS Groups and Activities
(It is not a complete proposal, just a draft)

Preparation/Introduction:
+ Visualization and analysis of dataset. :
      - Number of tweets, Users
      - Analysis of user names : Common words used, N-gram approach.
      - Geo-Spatial analysis if exists.
+ Converting everything to English.
      - Some tweets are in different language. Basically cleaning data.

Modeling:
+ Learning/Predicting model (SVM or something else).
+ Collecting features from Data.
+ Trying weighted SVM for weighted features (Weight words with higher chances of being malicious for terrorist activities ex. Bomb is more weighted than explosion)

Final Goal:
+ Predicting if a tweet is provocating or involves any terrorist activities.
+ Identifying recruitment activities of ISIS

Example test tweets : Its expected prediction
1) Data Science project presentation : False
2) Bombing in North Pole : True
3) Car explosion in South Pole : False
4) Baghdad tanks deployed : True
5) Syrian Troops Clash With Rebels as Bombs Fall on Aleppo : True

Some edge cases, I can think of: (Will be more while implementation)
+ Some words are not always related to Terrorist activities (ex. Tanks can be used for War or Terrorist activities) : Penalizing such words during training may fix the problem.
+ We need to find community or graph which can help us examine tweets better. Adding some ISIS Fan-boy in tweet and tweeting about science is not terrorist activity.
+ Updating ISIS tweets will help us classify future tweets better. Reason is, contain of a tweets can vary w.r.t current ISIS activities. Ex. Location and Type of attack (Chemical, Nuclear etc). This case is more of solving input of streaming tweets and classifying them as ISIS activities or not. (Passive aggressive algorithm can be used)
+ May require building a dictionary for identifying malicious tweets.

To be discovered
Naming conventions : Its very important, since we don't want to write tweets with terrorist activities and non terrorist activity all the time. Having names for each will help us and readers.
+ Tweets with terrorist activities named _____ and non _____.
+