# PROCESS FLOW

Cleaned Tweets

Data Stemming

Data Pre-Processing

Commercial

Non - Commercial

Relevant Tweets

Tweets Classification

Topic Modelling

Positive

Common Side effects- Tweet frequency

Negative

Disclosed and Non-Disclosed Adverse Drug Reaction

Supervised LDA

Non – Supervised LDA

Tweets and Twitter Handle details

Tweets Extraction- by drugname

* Twitter Extraction
  + By drug name
  + Getting tweets of 10 days old
  + Executing this step every week
* Identification of commercial and non-commercial tweets
  + Tweet characteristics
  + who posted - organisation or individual
  + who and what posted for the day
* Data Pre-processing and Data Stemming
  + Removal of stop words, punctuations, digits, symbols, etc
  + Removal of retweets
  + Converting each tweet into stemmed list of words
* Relevant tweets
  + Getting relevant tweets after matching the cleaned dataset with side-effect bag of words
* Tweets classification
  + SVM Classification into positive and negative tweets
  + Negative tweets - get the side effect frequency
  + Negative tweets - get the disclosed and undisclosed ADR
* Topic Modelling
  + On the cleaned tweets built an unsupervised topic model using Latent Dirichlet Allocation (LDA)
  + On the cleaned tweets built a supervised topic model using term frequency–inverse document frequency (TF-IDF)