

# Data Science Assignment

## Back ground

Monalisa, who is a Life sciences researcher does Pharmacovigilance (post-market drug safety surveillance) for Pharma Products.

Pharmacovigilance identifies ADRs (adverse drug reactions) after a drug's release. Comprehensive knowledge of ADRs can reduce the detrimental impact on patient's health system. Practically, clinical trials cannot investigate all settings in which a drug will be used, making it impossible to fully characterize the drug's adverse effect profile before its approval. Pharmacovigilance methods continuously analyse frequently updated data sources, Twitter in particular because of its large user base, demographic variability, and publicly available data.

ADR detection in social media requires automated methods to process the high data volume. It would greatly help her if she is able to automate the segmentation of Tweet into either ADR or NON-ADR, on the basis of Drug, Symptom and Effect mentioned in future.

Your Job is to Classify the Tweet into the 2 segments, please describe your approach and share the Python Code for the same.

## Data Set:

Attached data set (data.csv) has tweets labelled as ADR and NON-ADR. The file has information which contains column like user IDs, tweet IDs, and label along with the tweet.