

# AI & ML Video Assignment

5th October 2020

## Building a Web News Article Clustering System

### Latent Semantic Association

Perform topic modelling on the dataset using PLSA.

Dataset Link - <https://www.kaggle.com/therohk/india-headlines-news-dataset>

Output/explain the following steps in the code:

- a) The vocabulary of words used.
- b) Words document Co-occurrence matrix
- c) Words under each topic and the score

For the given dataset, use the first 100,000 headlines and 20 topics.