

Problem Statement:

Many news aggregators face the challenge of common news stories published from many providers. The importance of grouping / deduping is very much important, so it can save end users' reading time.

Dataset:

<https://www.kaggle.com/uciml/news-aggregator-dataset>

Above the kaggle dataset is a collection of news stories collected and grouped by similarity of topic being discussed.

Column "Story" is common for grouped articles.

Expected Outcome:

Develop algorithm technique to group articles by topic/story, and present accuracy analysis. Solution should be developed in python using machine learning library like scikit-learn, tensorflow, pytorch.

Submission:

Python scripts or jupyter-notebook should be shared along with requirement.txt