

## Data sources

### Scrapers

- Scrape content off social media (Facebook, LinkedIn, etc.....)
- Scrape posts based on selected criteria/keywords and groups

### Databases

- Store scraped content into databases (MongoDB / Firebase DB) to be acquired by the pipeline

## Pipelines

- Acquire data in an efficient way
  - Batching
  - Prefetching
- Includes preprocessing to required format based on the type of model trained
- Generate the data one item at a time to train the models

## Models

- Models and pipelines will eventually be hosted on platforms such as
  - Amazon SageMaker
  - Docker Image
- Models will be deployed on platforms such as
  - ACS
  - Firebase

## Application / Deployment

- Web based interface

### Front-end

- React Native

### Backend

- Firebase API

## Integration

- Integration of models, scraping and pipelines with the web app

### Frameworks used

- Nodejs

## Testing

- Unit tests for individual modules
- System tests for assessing integration between modules