Course Name: Solving Business Problems with NLP
Instructor: Juber Rahman

Project -1

Part A: Prepare a dataset for online games classification.

1. Scrape game reviews from commonsense media

   https://www.commonsensemedia.org/game-reviews.

   You may use ParseHub software for the scraping following this tutorial

   https://www.parsehub.com/blog/web-scraper-tutorial/

2. Manually label each row (game review) as safe or adult

Part B: Train an NLP model for classification

1. Split your data into train and test sets.
2. Preprocess the reviews for tokenization/ stop word removals
3. Prepare two set of embeddings using Bag-of-Words and TF-IDF
4. Train a machine learning model to classify the reviews into safe and adult
5. Evaluate your model on the test set.
6. Compare the performance for each embeddings
7. Upload your notebook in the course GitHub repo.

Reference: https://omdena.com/blog/internet-safety-children/