

Knowledge Discovery and Data Mining
Ton Duc Thang University, Ho Chi Minh City, Vietnam
-----oOo-----

Final Project

Task 1. (40%) Collecting data

1. Choosing three Vietnamese videos on YouTube with at least 100 comments. Crawling the first 100 comments of each videos.

Please find the Crawl Data Example.ipynb for an example of crawling.

2. Label these comments by yourself in seven classes: Disgust (chán ghét), Enjoyment (thích thú), Anger (giận dữ), Surprise (ngạc nhiên), Sadness (buồn bã), Fear (sợ hãi), Other (khác).

Task 2. (30%) Emotion Recognition for Vietnamese Social Media Text

1. Using training and validation sets of UIT-VSMEC (UIT-VSMEC.zip) for training your Emotion Recognition model (classification model).

Note: you can choose any technique in preprocessing, any classifier that you want.

2. Report the performance metrics (Accuracy, F1-score...) for the test set in UIT-VSMEC dataset.

Task 3. (30%) Applying the trained model in Task 2 to three datasets in Task 1. Then, report the performance metrics (Accuracy, F1-score...) for these datasets.

Team: two students for each group.

Submission:

- Online:

+ Form: https://docs.google.com/forms/d/e/1FAIpQLScIY6Xxe-_KdQOZFw_On8GkqLkfOZapbwUqaZgH7hheJlkDuQ/viewform

+ Submission file: python codes, presentation, project report.

+ File name: follows the rule in submission form.

+ Deadline: **23:59 PM May 16, 2021**

- Offline: project report (20-30 pages, print in two sides), submit on **the presentation date** (from May 17-23).