

# CS462: Natural Language processing

## Assignment 3

### Objectives

Practice how to perform text classification using a machine learning classification model and combinations of word embeddings or sentence embeddings as a feature vector

### Dataset

Movie reviews data set V2.0 contains 2000 text samples divided into 1000 positive reviews and 1000 negative reviews. Reference and download:

<http://www.cs.cornell.edu/people/pabo/movie-review-data>

### Task:

Load review samples (both positive and negative) and generate sentence embedding vector for the samples. You are required to use TWO of the embedding methods below:

- Learn word embeddings with Skip diagram **or** use CBOW.
- Build Sentence embedding based on the average of word embeddings
- Generate labels vector for the dataset. Randomly divide data to training and testing sets. Note that each set should contain samples of the two types
- Train a classification model to predict the label of the review.
- Train with **2** Models (SVM , CNN,...), each model train multiple times to optimize hyper parameters.
- Output

## CS462: Natural Language processing

### Assignment 3

- Report that includes comparison between tf-idf and skip diagram and CBOW, for the results of model (1) from assignment 2 and three models from assignment 3.
  - In the report mention model you used and best hyper parameter for each model
- 
- **Teams:** Form teams of 3 for this assignment same team as assignment 2.