

## **CENG463 Introduction to Machine Learning**

### **Assignment #2**

**Due date: 05.06.2021**

In this assignment, you're expected to visualize the outputs of two topic clustering methods on short texts using different dimensionality reduction algorithms.

- Use the StackOverflow short text dataset below:

<https://github.com/jacoxu/StackOverflow>

- Use LDA (Latent dirichlet allocation) and DMM (A dirichlet multinomial mixture model-based approach for short text clustering) as the clustering methods.
- Use PCA (Principal components analysis), MDS (Multi-dimensional scaling), LDA (Linear discriminant analysis) and t-SNE (Stochastic neighbour embedding) as the dimensionality reduction techniques.

Please give comment on the generated visualizations.