**CS5542 Big Data Apps and Analytics**

**In-class Exercise #3**

**Linear Regression:**

Implement Linear Regression Model on provided 3D Road Network (North Jutland, Denmark) Data Set.

1. Build the model using 50 iterations
2. Split data into 70% training and 30% testing
3. Report Training Mean Square Error
4. Report Test Mean Square Error

**K-Mean Clustering:**

**Implement K-Mean Clustering on provided 3D Road Network (North Jutland, Denmark) Data Set.**

1. Cluster the data into 3 and 4 classes
2. Evaluate clustering by computing Within Set Sum of Squared Errors
3. Compare their results

Dataset Link:

<https://umkc.box.com/s/82tetnw6ob9of7iuuchxj1p9u1srk95l>

**Reference:** [**https://archive.ics.uci.edu/ml/datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29**](https://archive.ics.uci.edu/ml/datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29)