**Lab 8: Topic Modeling**

By Sawit Koseeyaumporn

65070507238

In this Lab, we demonstrate to students how to apply topic modeling to real-world data.

Students will gain hands-on experience through this example.

* Download the datasets

A screenshot of a computer program

Description automatically generated

* Read the dataset and convert it dataFrame shape

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* Uses Gensim Library to convert a document into a list of tokens.

A screenshot of a computer

Description automatically generated

**Task 1: ID-to-word mapping**

A screenshot of a computer program

Description automatically generated

* Trying the different n\_topics from the LDA models and found out what is the words in each topic.

A screenshot of a computer

Description automatically generated

* Installing the pyLDAvis for visualization

A screenshot of a computer screen

Description automatically generated

* Using lda\_model, corpus and dictionary to create the visualization

A computer screen shot of a computer program

Description automatically generated

A screenshot of a graph

Description automatically generated

**Task 2: Try tuning the LDA parameters or incorporate additional text preprocessing**

* First, we will add additional text preprocessing

A screen shot of a computer program

Description automatically generated

A screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* Create the iteration to find the LDA parameter

A screen shot of a computer program

Description automatically generated

A computer screen with many white and green text

Description automatically generated

A black background with white text

Description automatically generated

สรุปว่า Num\_topics=2 และ num\_terms = 2000 เหมาะสมที่สุดสำหรับ Data นี้

* Visualize again

A screenshot of a computer

Description automatically generated

Sawit Koseeyaumporn 65070507238