**Please read the notes on configuration file and modify the config file in**

**..\NANDURI\_DSC550\_FINALPROJECT\dsc550\config\config.cfg**

**Due to the size of the source file, the original file is not being included and less number of features were considered for building models.**

**End to end pipeline:**  Performs end to end pipeline operations of building machine learning models and using those models to perform topic modelling.

File: wf\_endtoend\_model.py

Involves the below workflows

**Workflows**

workflow1\_prepdata: Workflow to prepare data and vectorize the data. Involves the below tasks defined in the following files:

task1\_preprocessing

task2\_vectorizer

workflow2\_buildevaluate\_models: Workflow to train and evaluate the model. Involves the below tasks defined in the following files:

task3\_model\_training

task4\_model\_evaluation

workflow3\_generatereport: Workflow to consolidate and generate the reports based on the trained models and their performance metrics. Involves the below tasks defined in the following files:

task5\_consolidate\_evaluation

task6\_generatereport

workflow4\_Modelling: Workflow to use the trained models and perform topic modelling on new text. Involves the below tasks defined in the following files:

task7\_vectorizer\_newtext

task8\_topic\_modelling\_vect

task9\_generate\_modellingrpt