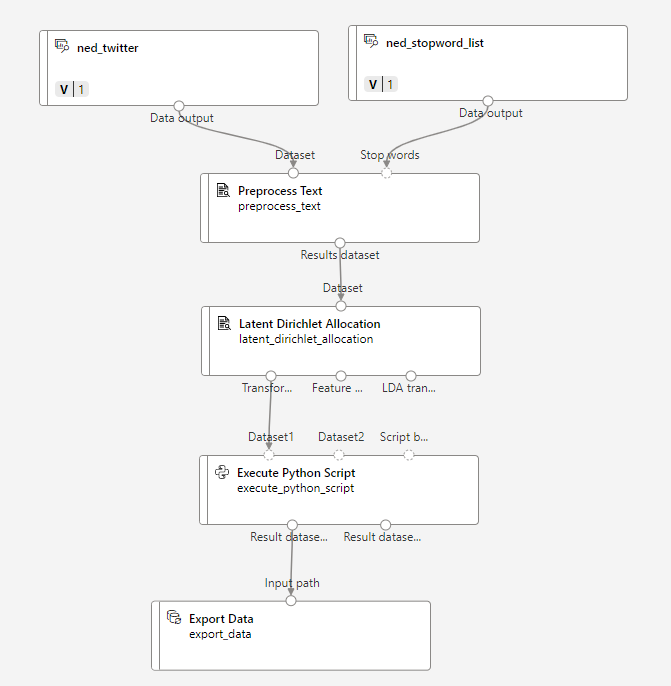
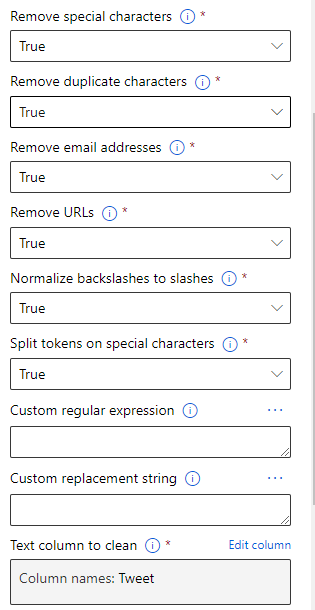
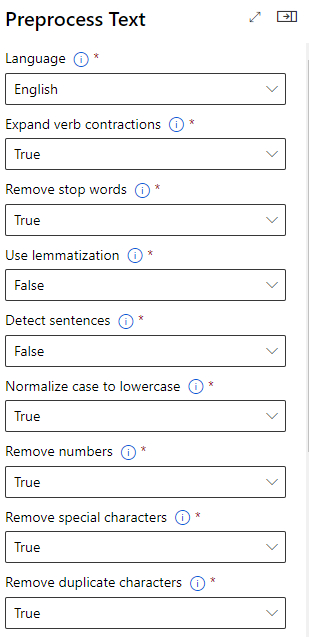
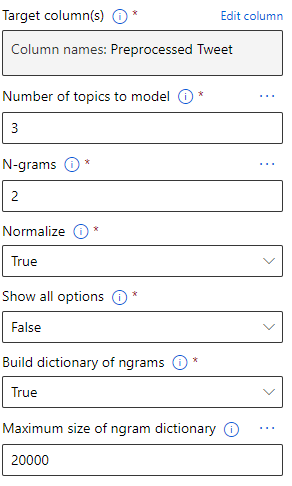
**Azure ML Designer Help sheet**

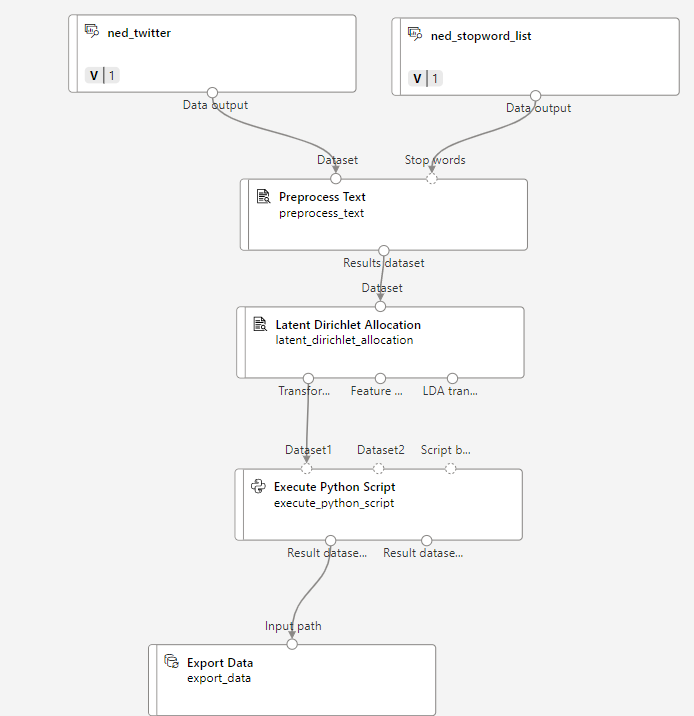
**LDA Topic Modelling**

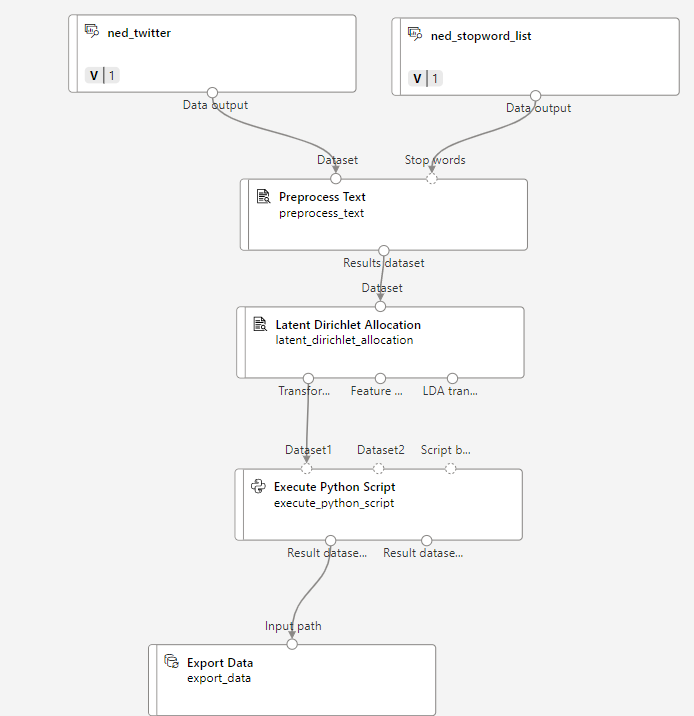
1. **Input data:** Create data assets in Azure ML studio for any data you want to incorporate in the pipeline.
   1. Create data asset for tweet dataset and link to next module
   2. Add a list of stopwords (also a data asset) and link to pre-process text module
2. **Preprocess text:** NLP techniques will require this step to run efficiently. Below are the possible options for this module.
   1. Remove stopwords was selected, utilising the stopword list added previously
   2. Remove URLs was useful in removing the article links within the tweets



1. **Latent Dirichlet Allocation (LDA):** Will apply topic modelling to a column in the dataset
   1. Specify pre-processed text column
   2. Number of topics was set to 3 for this scenario, however this can be adjusted for optimisation
   3. Connect transformed dataset to next module





1. **Execute Python Script:** Custom python script for formatting and assigning topic to each tweet dependent on highest probability.
2. **Export Data:** Export the resulting dataset to a specified location in blob storage