Summarization Task

Introduction:

The task is to perform an extractive summarization task to by extracting the most important sentences in the model

The task as mentioned is to perform extractive summarization which differs from abstractive summarization by the dataset and the architecture

* Extractive summarization uses an encoder in its architecture
* Abstractive summarization uses an encoder-decoder in its architecture

The data:

I downloaded the WikiHow dataset which is processed as an extractive summarization task

<https://drive.google.com/uc?id=1jyNdc0fhrriZXArV-9R2UJMHfdQaQV3D>

The tools:

I used the Transforersum tool to do this task which offers solutions for both extractive and abstractive summerization

<https://github.com/HHousen/transformersum.git>

The Process:

1. I Cloned the repo and installed the env variable so the packages don't conflict (the documentation advice)
2. I downloaded the wikihow dataset
3. Created folders for both data/logs/weights that will be needed to provide to the training command
4. I loaded the trained weights to the model and used it to inference for the test cases

Test Cases:

The test cases are 3 paragraphs that are extractive inferred to the determined number of sentences in the inference command

Challenges:

* Installing the repo was tricky and required a lot of time

- The offered models for extractive summarization are limited

References:

<https://github.com/HHousen/transformersum.git>

<https://transformersum.readthedocs.io/en/latest>