Solution

**Task 1: Data extraction and pre-processing.**

The data could be crawled from the Instagram posts as text and stored in Azure blob storge. Next, clean and pre-process the data by removing hashtags, links, mentions, tokenization, etc. using Spacy or NLTK.

**Task 2: Topic Discovery.**

The topics can be discovered using technique called Topic modelling like LDA, NMF, etc.

**Task 3: Implementation.**

The topic classification could be done in various ways such as normal machine learning classification models for example Decision trees, Naïve bayes, etc. Or we can fine tune the LLMs on the labelled dataset of posts for each topic. The dummy random assigning of score for the topic is attached.

The API end point is also attached and written using python flask.

For the deployment, we can make use of Azure function app or we can do locally as well as shown in the dummy code.

**Task 4: Estimations**

This solution (Dummy model and rough sketch) : 4 hours.

The complete application: 4 to 5 business days.

**Output:** This task I will be uploading it on Github repo.

**Task 5: Architecture design**

Instagram Data (Posts)

Azure Blob Storage

Data Pre-processing

Topic Discovery (LDA, NMF, Clustering)

Topic Classification (ML Classification, NLP many shot, or LLMs

Client app (Send POST request ex: Instagram post text)

Azure Function (Microservice endpoint)

Rest API Engpoint (Flask app)