Assignment 1

1. Establish connection to any social media platform using official APIs or available libraries. Use your own account credentials or API keys, retrieve **one sample** post/tweet/comment. (2 Marks)

- Mention the **API used**, **authentication mechanism**, and **rate limits**, if any.
- Highlight challenges encountered during API access, including limitations on free-tier usage.
- Save the sample data in JSON or CSV format.

You can refer to the following for the first question

https://wilpbitspilaniacin0.sharepoint.com/:b:/s/SocialMediaAnalyticsS2-24_AIMLCZG522/EVqb31DkcuFArjpBNHkOdksBhX2FEWnsGi05mYVe_OdwOw?e=qelfqn

2. Perform sentiment analysis on a dataset. The dataset includes review and sentiment (positive or negative). The objective is to analyze the sentiments expressed in reviews, develop a sentiment classifier, and gain insights

1. https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews
2. https://www.kaggle.com/datasets/bittlingmayer/amazonreviews
3. https://huggingface.co/datasets/cornell-movie-review-data/rotten_tomatoes
4. https://www.kaggle.com/datasets/jiashenliu/515k-hotel-reviews-data-in-europe
5. https://www.kaggle.com/datasets/therohk/million-headlines

Data set 1 : Groups 1- 12

Data set 2 : Groups 13- 24

Data set 3 : Groups 25- 36

Data set 4 : Groups 37- 48

Data set 5 : Groups 49 - 60

Instructions:

1. Data Preparation and Exploratory Analysis (2 Marks):

- Preprocess the text data, including tasks such as tokenization, lowercasing, and removal of stop words and special characters.

- Analyze the distribution of sentiment labels in the dataset.

- Explore the most frequent words associated with each sentiment class.

- Visualize sentiment distribution

2. Feature Extraction and Model tuning (2 Marks):

- Utilize appropriate feature extraction techniques such as TF-IDF or word embeddings. (2 Marks)

- Train a machine learning algorithm (e.g., Naive Bayes, Support Vector Machine) or a deep learning model (e.g., LSTM, BERT).

3. Model Evaluation and Compare with any LLM of your choice (2 Marks):

- Evaluate the model's performance using accuracy, precision, recall, and F1 score metrics.

You have to submit both the code and a pdf document that contains the output. If you miss any the assignment will not be graded.