DATASET REPORT

**Week – 1 Task Report : Dataset overview**

**Enrolment Number : 21CS002362**

**Name : Bairagoni Manikanth**

**Branch : CSE (AI & ML) [Batch 2]**

**Mentor : Sevy**

**Dataset Overview :**

The dataset consists of 18,409 comments collected from 1,869 unique YouTube videos. Each entry in the dataset includes the video ID, the text of the comment, the number of likes the comment received, and a sentiment score indicating whether the comment is positive, neutral, or negative.

The distribution of likes across comments varies significantly, with a mean of 1,040 likes per comment and a maximum of 891,372 likes. The median number of likes is 29, indicating that while some comments gain substantial popularity, most receive a modest amount of likes.

In terms of sentiment, the dataset reveals that the majority of comments (62.10%) are positive, reflecting a generally favorable response. Neutral comments account for 25.20% of the dataset, while negative comments make up the smallest portion at 12.70%. This distribution suggests that the overall sentiment is predominantly positive, with a smaller but notable presence of neutral and negative sentiments.

This comprehensive dataset provides a valuable basis for analyzing public opinion and engagement on YouTube, offering insights into the sentiment trends and the popularity of comments across various videos.

**Column Descriptions**

1. **Unnamed: 0**
   * **Description:** An automatically generated index column. It serves as a unique identifier for each row in the dataset.
   * **Type:** Integer
2. **Video ID**
   * **Description:** A unique identifier for the YouTube video associated with each comment. This ID allows us to trace comments back to their original video source.
   * **Type:** String
3. **Comment**
   * **Description:** The text content of the comment made by a user on a YouTube video. This field contains the actual written feedback or opinion expressed by the user.
   * **Type:** String
4. **Likes**
   * **Description:** The number of likes that the comment received. This metric indicates the level of agreement or popularity of the comment among other viewers.
   * **Type:** Integer
5. **Sentiment**
   * **Description:** The sentiment score assigned to each comment. This score reflects the emotional tone of the comment and is categorized into three types:
     + **0:** Negative sentiment
     + **1:** Neutral sentiment
     + **2:** Positive sentiment
   * **Type:** Integer

**Summary Statistics:**

* **Total Number of Comments:** 18,409
* **Number of Unique Videos:** 1,869

**Likes:**

* **Mean:** 1,040.02
* **Standard Deviation:** 10,651.37
* **Minimum:** 0
* **25th Percentile:** 5
* **Median:** 29
* **75th Percentile:** 190
* **Maximum:** 891,372

**Sentiment Distribution:**

* **Positive (2):** 11,432 comments (62.10%)
* **Neutral (1):** 4,639 comments (25.20%)
* **Negative (0):** 2,338 comments (12.70%)

**Key Observations:**

* The dataset has a predominant positive sentiment, with over 60% of the comments classified as positive.
* Neutral comments account for about a quarter of the dataset, and negative comments are the smallest group, comprising around 13%.
* The likes distribution shows a wide range, indicating that while a few comments are extremely popular, the majority receive a moderate number of likes.

This dataset provides a rich source of information for analyzing user sentiment and engagement on YouTube, allowing for detailed insights into public opinion and the popularity of user-generated content.

**Concluding Remarks:**

The dataset consists of 18,409 comments collected from 1,869 unique YouTube videos, with each comment containing attributes such as the video ID, comment text, number of likes, and a sentiment score. The sentiment distribution reveals that 62.10% of the comments are positive, 25.20% are neutral, and 12.70% are negative, indicating a predominantly favorable sentiment. The number of likes per comment varies widely, with an average of 1,040.02 likes, a median of 29 likes, and a maximum of 891,372 likes, highlighting that while some comments are extremely popular, most receive a moderate number of likes. Overall, the dataset provides a detailed view of user sentiment and engagement on YouTube, offering valuable insights into public opinion and the popularity of user-generated content.