**Name Aoun Muhammad**

**Roll no 045**

**Project Sentiment analysis youtube coment**

**Submit Sir Muhammad Rasikh Ali**

**Project**

**Sentiment Analysis on YouTube Comments**

**Introduction**

The goal of this project was to analyze the sentiment of a sample of YouTube comments using Python. Sentiment analysis, a key aspect of Natural Language Processing (NLP), provides insights into the emotional tone of user-generated content, helping content creators understand audience perceptions.

This project leverages libraries such as **Pandas**, **re** (for text preprocessing), **TextBlob** (for sentiment analysis), **Seaborn**, and **Matplotlib** (for visualization).

**Objectives**

1. **Clean YouTube Comments:** Remove unnecessary characters, URLs, and standardize the text to lowercase.
2. **Perform Sentiment Analysis:** Classify comments into **Positive**, **Neutral**, or **Negative** categories.
3. **Visualize Sentiment Distribution:** Create a bar chart showing the frequency of each sentiment category.

**Dataset**

**Input Data:**

* A sample of seven YouTube comments, ranging from positive to negative sentiments. Some comments included URLs or special characters.

**Sample Comments:**

1. "I love this video!"
2. "bad content"
3. "amazing video quality"
4. "Amazing effort! Keep up the great work!"
5. "I am not like the audio quality."
6. "this is the best video I have seen"
7. "Check out this amazing content: <https://youtu.be/Gx5qb1uHss4?si=CcFl6YwxH0lhl1Ie>"

**Methodology**

1. **Preprocessing Comments:**
   * Removed URLs using the re.sub() function.
   * Stripped non-alphabetic characters (e.g., punctuation).
   * Converted all text to lowercase for uniformity.
2. **Sentiment Detection:**
   * Used the TextBlob library to calculate the polarity of comments.
     + **Positive:** Polarity > 0
     + **Neutral:** Polarity = 0
     + **Negative:** Polarity < 0
3. **Visualization:**
   * Utilized the **Seaborn** library to create a bar plot representing the frequency of sentiments.

**Results**

**Sentiment Analysis Summary:**

| **Comment** | **Sentiment** |
| --- | --- |
| I love this video! | Positive |
| bad content | Negative |
| amazing video quality | Positive |
| Amazing effort! Keep up the great work! | Positive |
| I am not like the audio quality. | Negative |
| this is the best video I have seen | Positive |
| Check out this amazing content: <https://youtu.be/Gx5qb1uHss4?si=CcFl6YwxH0lhl1Ie> | Positive |

**Visualization:**

* The generated bar chart showed a higher frequency of **Positive** sentiments compared to **Negative** and **Neutral**.

**Conclusion**

This project demonstrated a simple yet effective sentiment analysis pipeline for YouTube comments. The key insights are:

* Most comments were **Positive**, reflecting a generally favorable audience reaction.
* Preprocessing significantly improved text analysis accuracy by removing irrelevant data like URLs.
* Sentiment analysis tools like **TextBlob** can efficiently classify short texts.

**Limitations & Future Work**

1. **Limitations:**
   * The dataset was small, which might not represent a wider audience's sentiment.
   * TextBlob has limitations in understanding complex sentiment expressions (e.g., sarcasm).
2. **Future Enhancements:**
   * Use a larger dataset with varied comments for better insights.
   * Explore more advanced NLP techniques (e.g., BERT or VADER sentiment analysis).
   * Include multilingual comments for broader applicability.

