**Lab 1**

**Social Media Sentiment Analysis**

**Objective**: To analyze sentiments (positive, negative, neutral) in social media posts.

**Key Concepts**: Natural Language Processing (NLP), Sentiment Analysis.

**Tools and Technologies**

**Libraries**:

* + ***tweepy*** for accessing the Twitter API (if using Twitter data).
  + NLP libraries like ***NLTK*** or ***TextBlob*** for sentiment analysis.
  + ***pandas*** for data manipulation.
  + ***matplotlib*** or ***seaborn*** for visualization.

**Dataset**

* **Twitter API**: Use ***Tweepy*** to collect tweets based on specific keywords, hashtags, or user accounts. Ensure compliance with Twitter's data usage policies.
* **Reddit API**: Alternatively, use Reddit data focusing on specific subreddits or topics.
* **Pre-existing Datasets**: There are also pre-existing datasets available for sentiment analysis that can be used for this project: <https://www.kaggle.com/datasets/kazanova/sentiment14> , <https://www.kaggle.com/c/si650winter11>

**Tasks Breakdown**

1. **Data Collection**

* Use social media APIs to collect data. This could involve streaming live tweets, scraping historical posts, etc.
* Store the collected data in a structured format (like a CSV file or a database).

2. **Data Preprocessing**

* Cleaning Text Data: Remove noise (e.g., URLs, mentions, hashtags, special characters).
* Normalization: Convert text to lower case, remove stop words, perform stemming/lemmatization.
* Tokenization: Break text into words or tokens.

3. **Sentiment Analysis**

* Use libraries like **NLTK** or **TextBlob** to assign sentiment scores to posts.
* Classify sentiments as positive, negative, or neutral based on the scores.

4. **Exploratory Data Analysis (EDA)**

* Analyze the distribution of sentiments across different topics, timeframes, or demographics.
* Identify patterns or trends in sentiment in relation to external events.

5. **Visualization**

* Create visualizations to represent sentiment distributions (e.g., bar charts, pie charts, time series plots).
* Use word clouds to represent frequently occurring words in different sentiment categories

**IMPLEMENTATION**

1. Let`s create the project and download all necessary libraries:

**Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание**

pip install tweepy

pip install nltk  
pip install textblob

pip install pandas

pip install matplotlib

1. Let`s connect Twitter API