Problem Statement: Sentiment Analysis Using TF-IDF Vectors

Objective:

Design and implement a machine learning model to perform sentiment analysis on a given dataset containing textual data and corresponding sentiment labels. The analysis should use **TF-IDF (Term Frequency-Inverse Document Frequency)** vectors to transform the text into numerical features and classify the sentiments effectively.

[Dataset Description for Sentiment Analysis Project](#)

Content:

The dataset is stored in a **CSV file** and includes six fields:

1. **Polarity (Column 0)**:
   - Sentiment label for the tweet:
     - `0`: Negative sentiment
     - `2`: Neutral sentiment
     - `4`: Positive sentiment

2. **Tweet ID (Column 1)**:
   - A unique identifier for each tweet.

3. **Date (Column 2)**:
   - The timestamp of when the tweet was posted, in the format `Day Month Date HH:MM:SS UTC Year` (e.g., `Sat May 16 23:58:44 UTC 2009`).

4. **Query (Column 3)**:
   - The search query used to retrieve the tweet. If no query was used, the value is `NO_QUERY`.

5. **User (Column 4)**:
   - The username of the account that posted the tweet (e.g., `robotickilldozr`).

6. **Text (Column 5)**:
   - The content of the tweet, consisting of raw text after emoticons have been removed (e.g., `Lyx is cool`).