Assignment – 16

You are tasked with developing a Python code for sentiment extraction utilizing a provided sample dataset. The dataset consists of textual data annotated with labels categorizing sentiments into four categories: "rude," "normal," "insult," and "sarcasm."

Dataset:

- Real News: https://drive.google.com/file/d/1FL2HqgLDAP5550nd1h_8iBhAV-ISTnzr/view?usp=sharing
- Fake News: https://drive.google.com/file/d/1EdI_HyUeI_Fi2nld7rQnnGEpQqn_BwM-/view?usp=sharing

1. Outline the key steps involved in developing a sentiment extraction algorithm using Python.
2. Describe the structure and format of the sample dataset required for sentiment extraction.
3. Implement the Python code to read and preprocess the sample dataset for sentiment analysis. Ensure that the code correctly handles text data and labels.
4. Discuss the process of classifying sentiments into the specified categories: "rude," "normal," "insult," and "sarcasm." Explain any techniques or algorithms employed for this classification task.
5. Evaluate the effectiveness of the sentiment extraction algorithm on the provided sample dataset. Consider metrics such as accuracy, precision, recall, and F1-score.
6. Propose potential enhancements or modifications to improve the performance of the sentiment extraction algorithm. Justify your recommendations.
7. Reflect on the ethical considerations associated with sentiment analysis, particularly regarding privacy, bias, and potential misuse of extracted sentiments.
8. Write a complete code for this assignment.