

Reflection

For my project, the most important evaluation metric is the **F1-score**.

Since this task involves **sentiment analysis**, both **precision** (how many predicted positives are actually positive) and **recall** (how many actual positives were correctly identified) are crucial. A model that only focuses on accuracy might perform well overall but still misclassify many minority or subtle sentiment cases.

The **F1-score** provides a balanced view by combining precision and recall into a single measure. It ensures that the model performs consistently across both positive and negative sentiment classes, avoiding bias toward the majority class.

Therefore, the F1-score gives a more reliable indication of how well the model will perform in real-world scenarios where data is often imbalanced.