

Screenshot of my code's results:

F1 Score: 0.5936

Precision: 0.6806

Recall: 0.6197

Accuracy: 0.6197

Annotation complete. Results saved to annotated_sentences.csv

My code resulted in a F1 Score of 0.5936, Precision of 0.6806, Recall of 0.6197, and Accuracy of 0.6197. I have noticed that changes in the prompt gave differences in overall analysis of the sentence. It had a tendency of detailed prompts giving accurate analysis on quite “obvious” sentences, and loose detailed prompts catching the “underlying nuance” of a sentence. Along the given csv file’s labels, I categorized a few sentences based on my opinion just like how I did in Label Studio for a new comparison to the results given by the model, and within the interpretation this trait was more distinguished. I believe the difference in the prompt’s detail is within a similar context to overfitting and underfitting in classification models.