Flan-T5 is an open-source large language model (LLM) developed by Google. Flan-T5 is designed for natural language processing (NLP) tasks through instruction fine-tuning, which is a technique where a pre-trained LLM is further trained on a dataset consisting of instruction-response pairs. This allows it to better understand and follow human instructions. Flan has an encoder-decoder (Seq2Seq) architecture, which means its neural network is designed for sequence-to-sequence tasks where an input sequence is mapped to an output sequence. These characteristics make Flan-T5 a good algorithm to classify our Rotten Tomatoes dataset. As we will see, Flan-T5 had the highest performance metrics compared to the other LLMs tested. However, it does not perform as well as other LLMs (like ChatGPT) for conversational uses and creative text generation.

Another subject of interest for this project is prompt engineering, which is a technique of carefully crafting the prompts to guide the LLM towards a desired output. In this project a default prompt (prompt 1) was used with each LLM; however, to increase the prediction metrics the following prompts were tested:

* **Prompt 1 (Default):** Is the following movie review positive or negative?
* **Prompt 2:** Is the following movie review from rotten tomatoes positive or negative?
* **Prompt 3:** Pretend you are a cinephile. Is the following movie review from rotten tomatoes positive or negative?
* **Prompt 4:** Pretend you are a highly educated movie critic who likes dramas and documentaries. Is the following movie review from rotten tomatoes positive or negative?

While the Flan-T5 LLM yielded high prediction metrics (e.g. 85% accuracy), Table 1 shows that the different prompts resulted in little to no effect on the classification metrics. This could be because the LLM is overly sensitive to the sentimental classification task, causing it to essentially ignore the additional context added. It could also be because the prompts are simply not providing enough additional context.



**Table 1**: classification report for the different prompts tested.