This pipeline performs sentiment analysis using Hugging Face's Transformers library with a pre-trained BERT model. To keep it beginner-friendly and reduce computation time, we use a smaller portion of the IMDb dataset, which contains movie reviews labeled as either positive or negative.

The first step is preprocessing. We use BERT’s tokenizer to convert the raw text into numerical format that the model can understand. The tokenizer also handles padding and truncation, so all inputs are the same length.

Finally, we fine-tune the model using the bert-base-uncased version, which is a standard and widely used pre-trained BERT model.