

# Project Milestone

## Goals Defined for the Milestone:

- Select a publicly available dataset and preprocess it for sentiment analysis.
- Train a Transformer model on the data.
- Achieve a baseline accuracy of 70% on the test dataset.
- Implement a working script for dataset loading, preprocessing, and model evaluation.

## Challenges blocked the Progress:

- Training speed: The initial Transformer model (BERT) was too slow, taking over 7 hours. This was resolved by switching to the DistilBERT model.
- Limited compute resources: Training on larger datasets is challenging with the available hardware.

The screenshot shows a Jupyter Notebook cell with the following code and output:

```
# Train the model
trainer.train()
```

wandb: WARNING The `run\_name` is currently set to the same value as `TrainingArguments`  
wandb: Using wandb-core as the SDK backend. Please refer to <https://wandb.me/wandb-co>  
wandb: Currently logged in as: ml\_monster. Use `wandb login --relogin` to force relogi  
Tracking run with wandb version 0.18.7  
Run data is saved locally in /content/wandb/run-20241201\_214715-ztggqncf  
Syncing run [.results](#) to Weights & Biases (docs)  
View project at [https://wandb.ai/ml\\_monster/huggingface](https://wandb.ai/ml_monster/huggingface)  
View run at [https://wandb.ai/ml\\_monster/huggingface/runs/ztggqncf](https://wandb.ai/ml_monster/huggingface/runs/ztggqncf)  
[ 12/9375 12:11 < 190:12:38, 0.01 it/s, Epoch 0.00/3]  
Epoch Training Loss Validation Loss

To the right of the notebook cell is a sidebar showing system monitoring information:

- At your current usage level, this runtime may last up to 84 hours 10 minutes.
- Manage sessions
- Want more memory and disk space? [Upgrade to Colab Pro](#)
- Python 3 Google Compute Engine backend  
Showing resources from 1:46 PM to 2:02 PM
- System RAM: 10.2 / 12.7 GB (Graph showing usage)
- Disk: 33.7 / 107.7 GB (Graph showing usage)

## Next Steps:

- Tune hyperparameters to further improve the model's accuracy.
- Extend the training to a larger subset if compute resources allow.
- Implement and evaluate additional Transformer architectures.

Project Repo Link: <https://github.com/nihal2704/CS507>