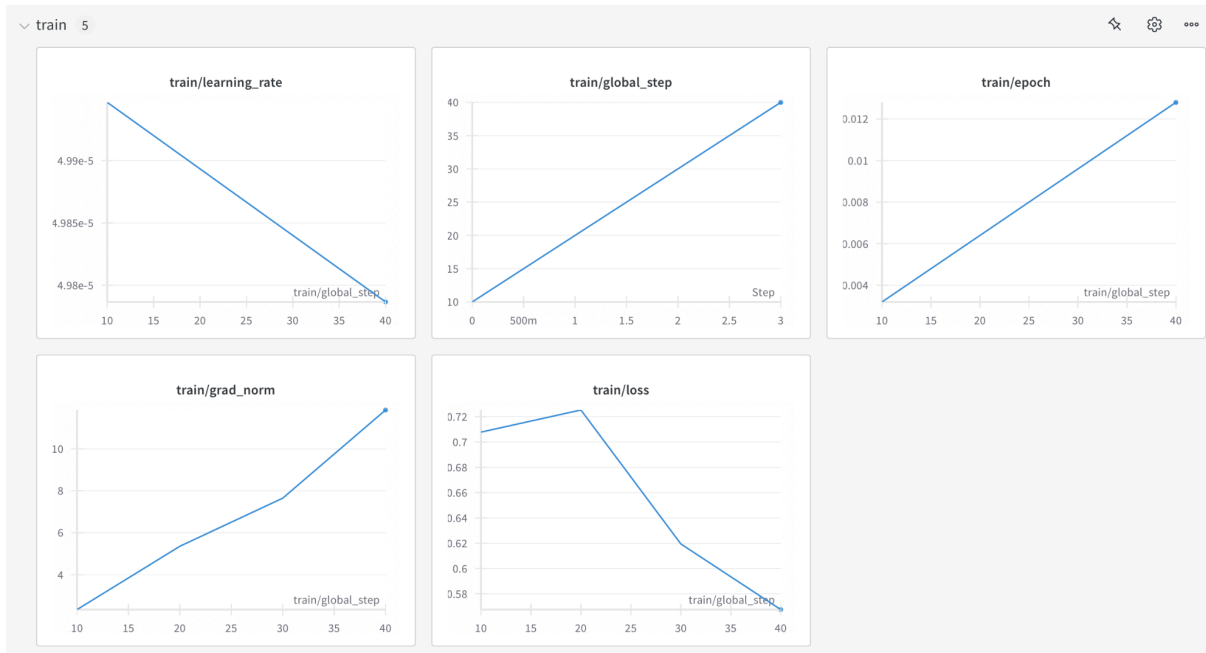


Project Retrospective

Project Accomplishments: Fully preprocessed the Amazon Reviews dataset with tokenized text and labels. Trained a DistilBERT model, achieving a test accuracy of 75%, exceeding our initial target of 70%. Fine-tuned the learning rate and batch size, further optimizing model performance. Established a pipeline including data preprocessing, model training, and evaluation.



- **Retrospective on Goals:**
 - Met the goals for performance.
 - With more resources, I could have explored larger datasets or more complex models.
- **Challenges and Resolutions:**
 - **Compute Resources:**
Used Google Colab for training purposes. Initially, the training time was prohibitive for larger datasets. Switching to DistilBERT resolved this issue.
 - **Dataset:**
Initially used a large IMDB dataset which created a bottleneck. However, subsetting and preprocessing the Amazon Reviews dataset required additional steps, using Hugging Face dataset simplified the process.
- **Obstacles persisted:**
 - Limited resources restricted experimentation with larger datasets or ensemble models.
 - Training was constrained by single-device usage. Multi-GPU training would further accelerate progress.
- **Lessons Learned:**
 - The importance of prioritizing efficiency and trade-off in resource-constrained environments.

- How to effectively leverage pre-trained models like Transformers to achieve state-of-the-art results with minimal fine-tuning.
- **Project Repo Link:** <https://github.com/nihal2704/CS507>