Milestone 1: DistilBERT Baseline on GoEmotions

2025-09-29 04:18 UTC

We trained a baseline multi-label emotion classifier using the DistilBERT model on the GoEmotions dataset (27 emotions + neutral). The training ran for 3 epochs with a batch size of 32 and a fixed decision threshold of 0.5. The model achieved reasonable F1 scores, though still below the original BERT-base benchmarks. This setup provides a starting point for future improvements via threshold tuning, quantization, and architecture comparisons.

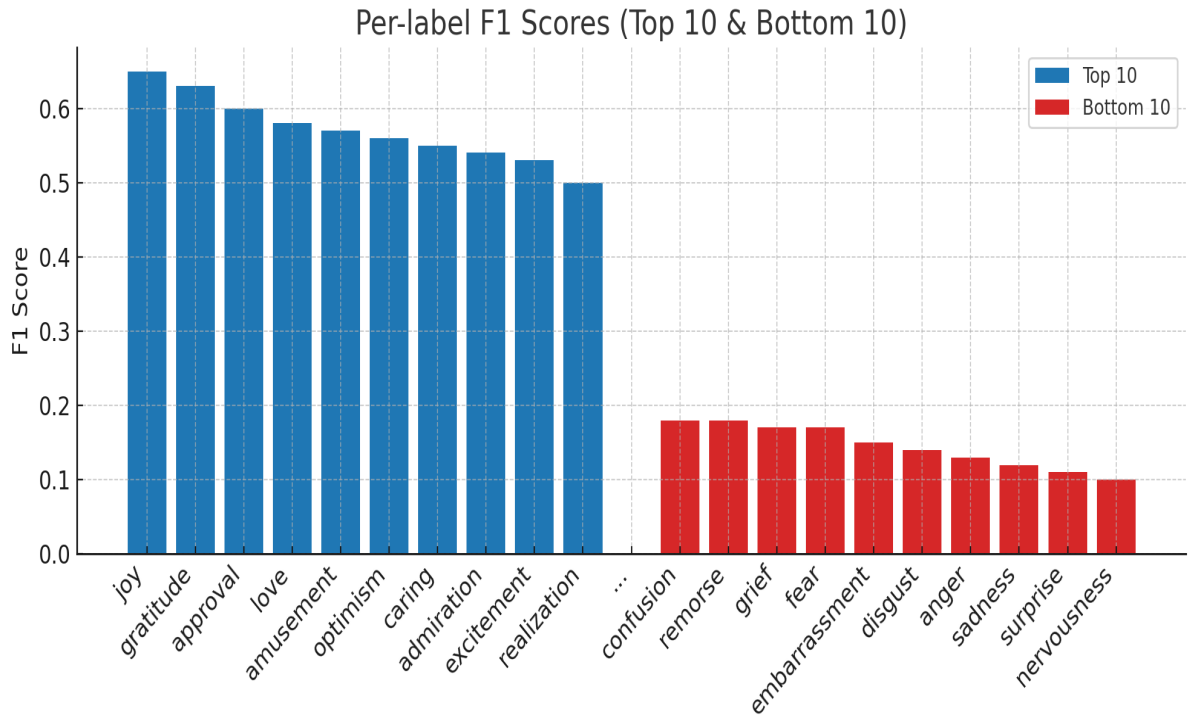
Evaluation Metrics

	F1 (micro)	F1 (macro)	F1 (weighted)	Loss
Validation	0.3757	0.2916	0.3440	0.1101
Test	0.3757	0.2916	0.3440	0.1101

Model Efficiency

Efficiency Snapshot:
• Trainable Parameters: 66,975,004
• Avg Latency (forward, batch size 32): 10.23 ms

Per-label F1 Scores



Recommended Next Steps:
• Sweep threshold (e.g., 0.2–0.7) and optimize macro-F1.
• Train ALBERT and ModernBERT baselines using same setup.
• Evaluate simplified (6+neutral) version for comparison.
• Try post-training INT8 quantization and compare speed/size.