

MuRIL Multilingual Classification Report

Generated on: 22 Dec 2025, 15:22

Model Information

Model: MuRIL (Multilingual Indian Language Representation)
Task: Multilingual Hate-Speech Classification
Split: 70% Train / 10% Validation / 30% Test

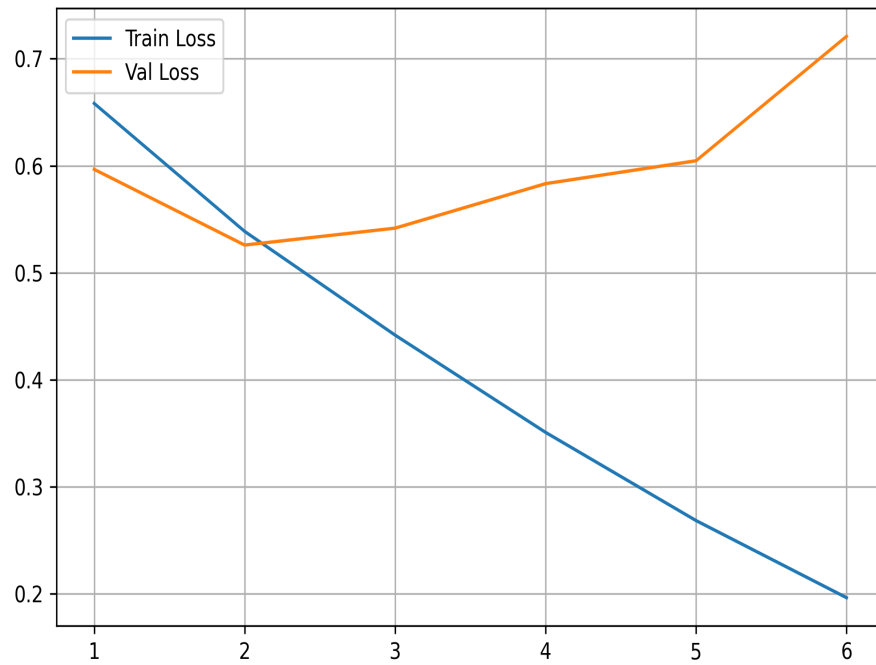
Training Hyperparameters

Parameter	Value
Learning Rate	2e-05
Batch Size	32
Epochs	8
Optimizer	AdamW
Scheduler	Linear + Warmup
Max Sequence Length	128
Early Stopping	Patience = 2 (Val F1)

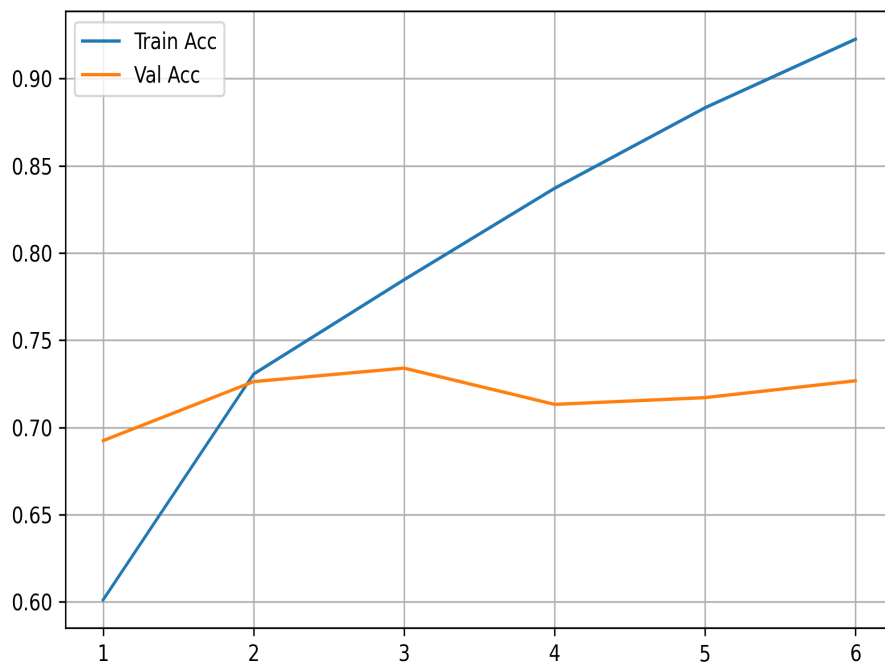
Test Set Metrics

Metric	Value
Accuracy	0.7166
Balanced Accuracy	0.7258
Precision	0.6476
Recall	0.8552
Specificity	0.5965
F1-score	0.7371
AUC-ROC	0.8240

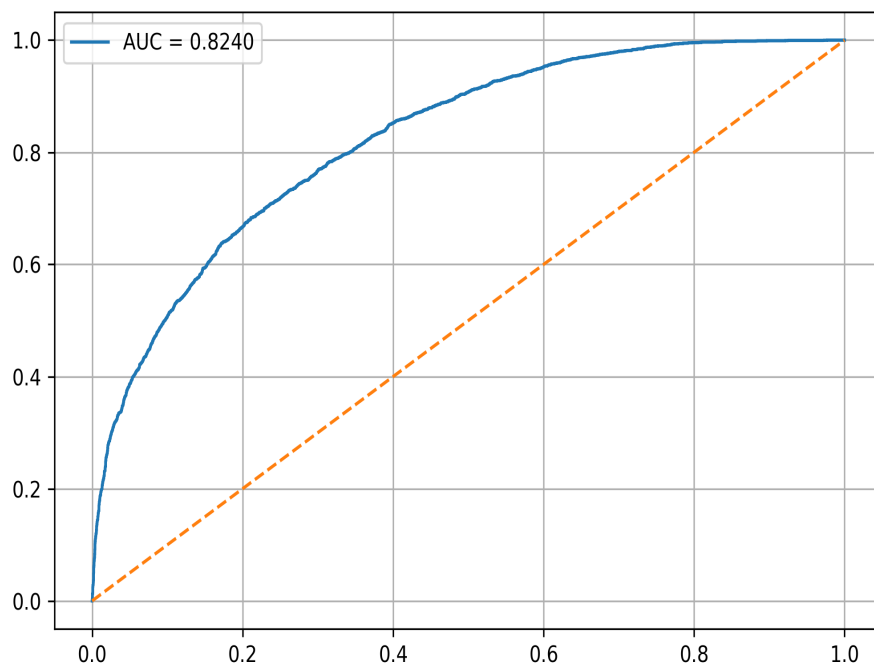
Train vs Validation Loss



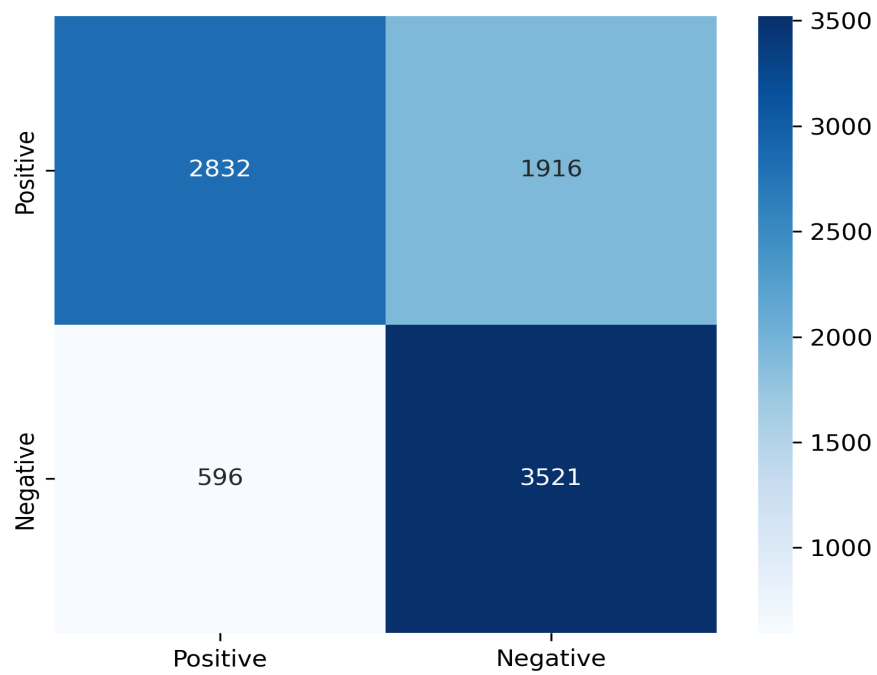
Train vs Validation Accuracy



AUC-ROC Curve



Confusion Matrix



Note: Best model selected using validation F1-score with early stopping. Final results are reported on the unseen test set.