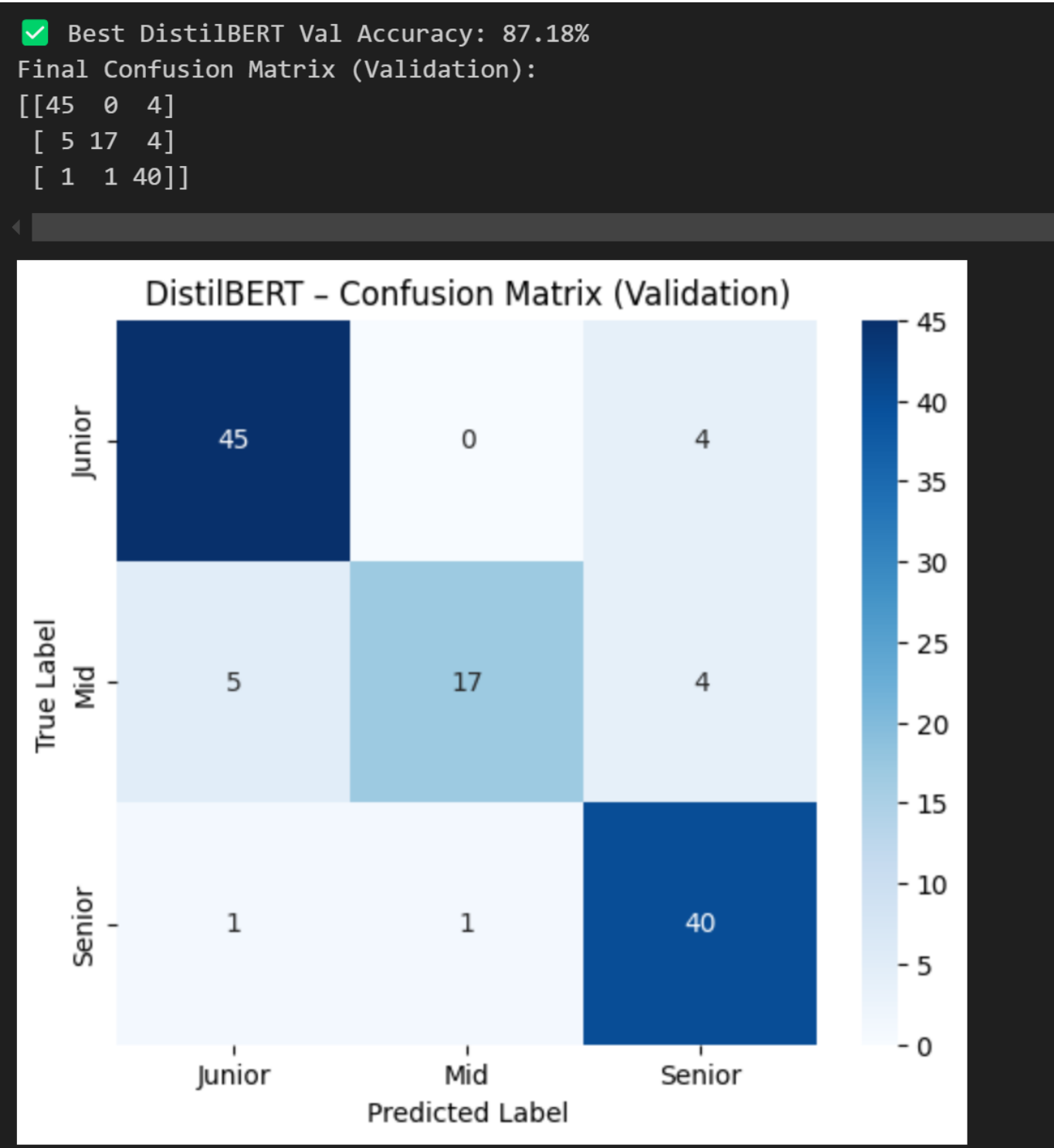


DistilBERT Classification Results

Best Validation Accuracy: 87.18%



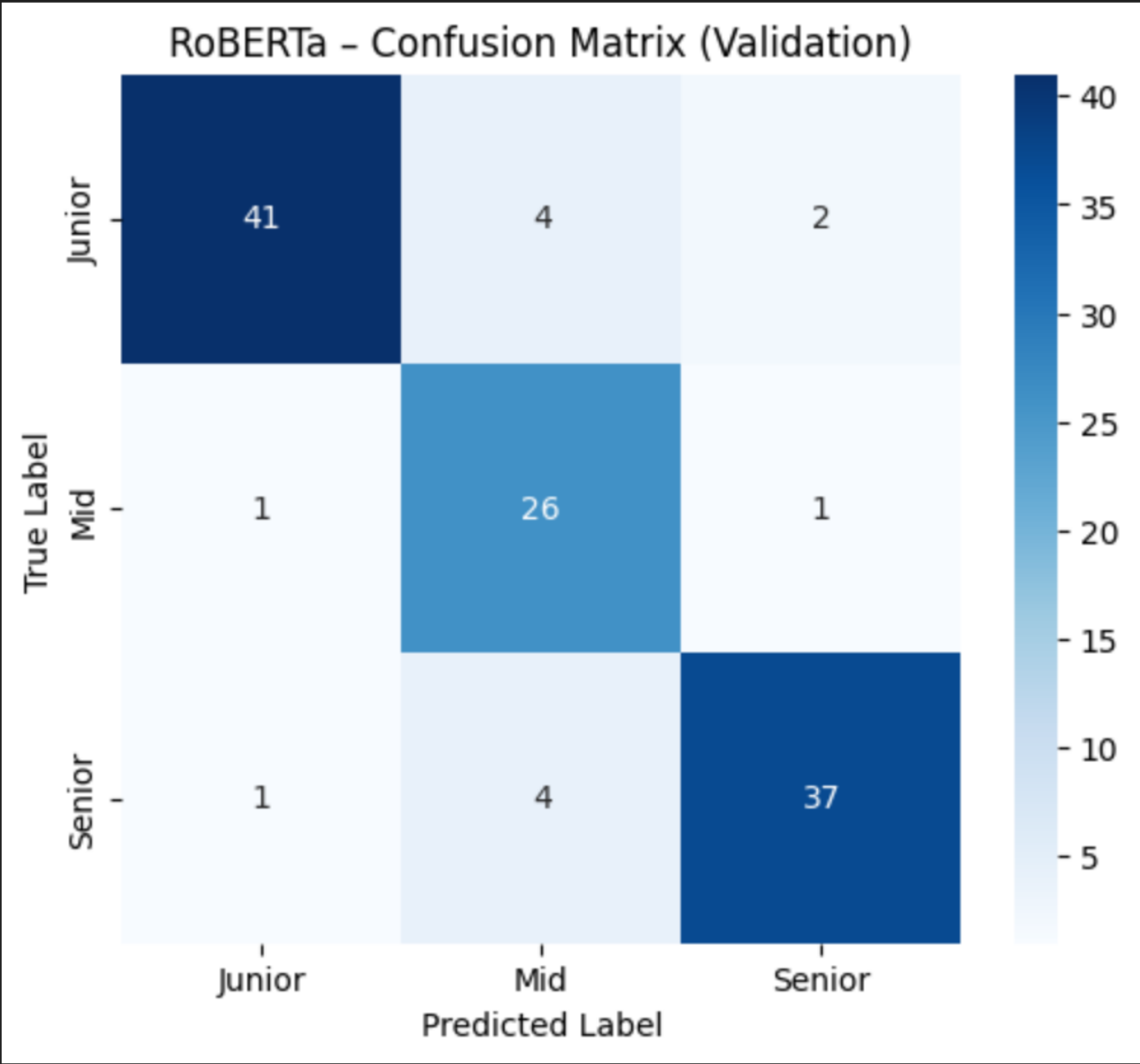
RoBERTa Classification Results

Best Validation Accuracy: 90.60%

✓ Best RoBERTa Val Accuracy: 90.60%

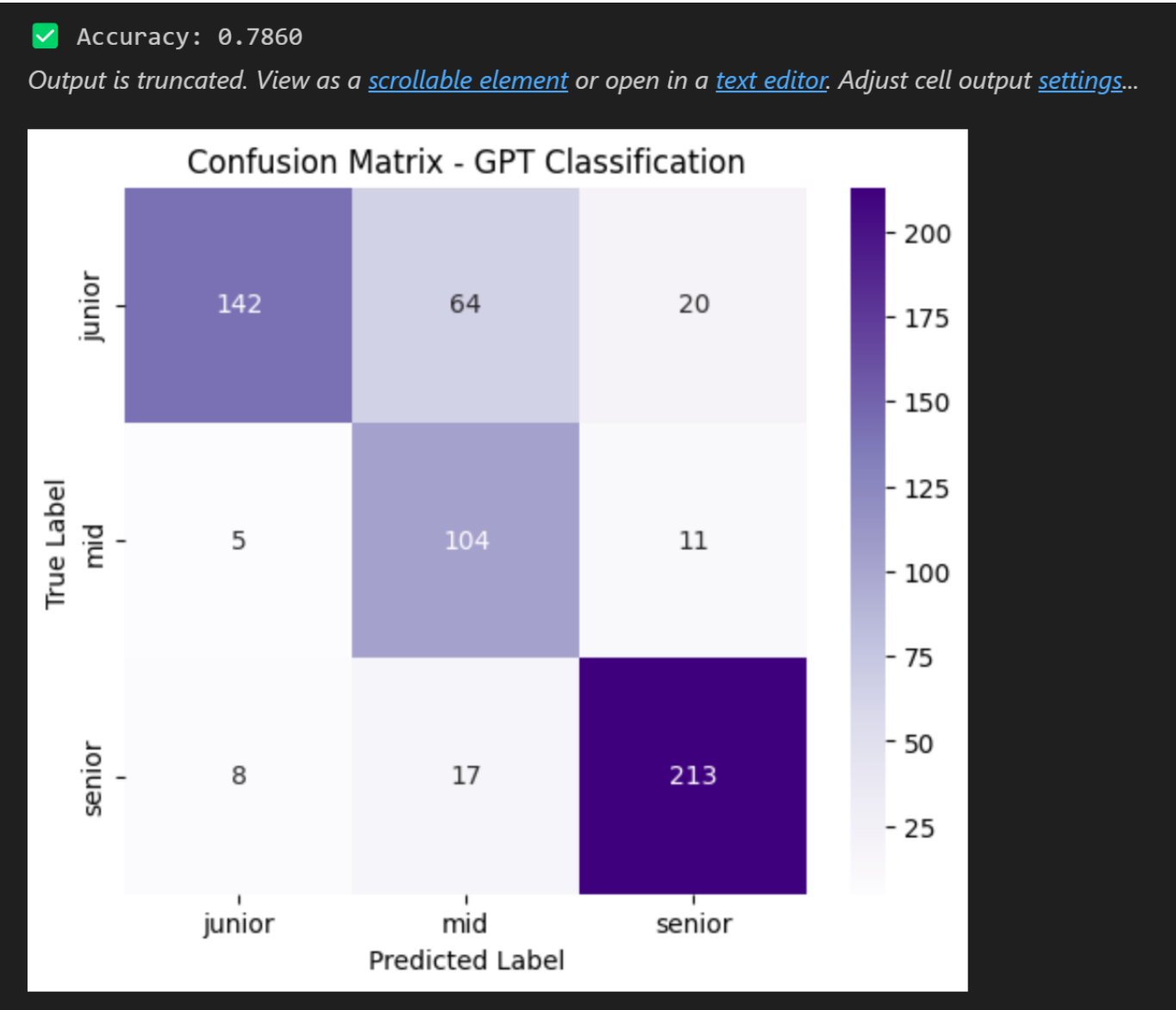
Final Confusion Matrix (Validation):

```
[[41  4  2]
 [ 1 26  1]
 [ 1  4 37]]
```



GPT-4 Classification Results

Accuracy: 78.60%



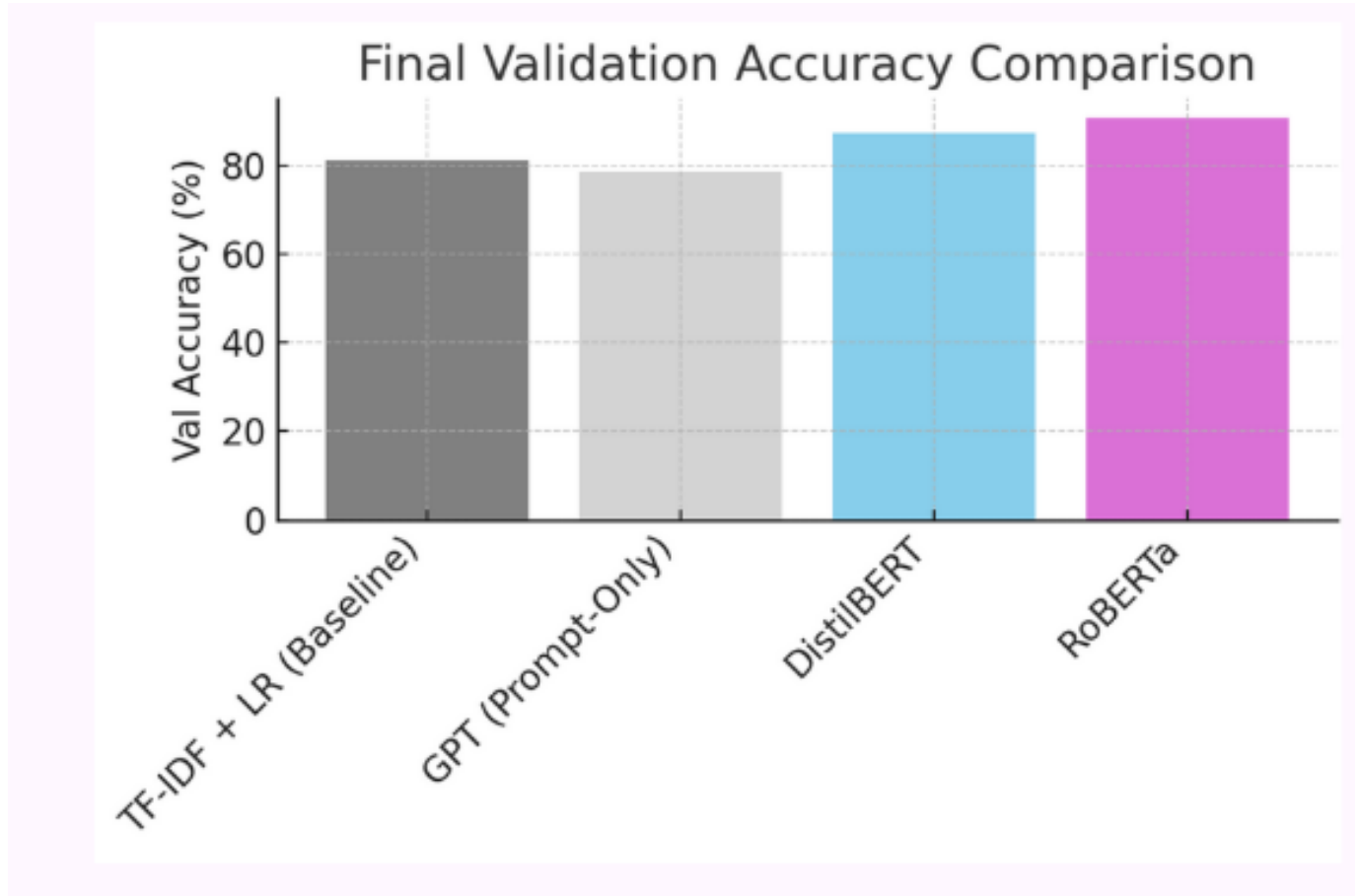
Model Performance Summary Table

Model	Best Epoch	Val Loss	Val Acc (%)
TF-IDF + LR (Baseline)	-	-	81.20
GPT (Prompt-Only)	-	-	78.60
DistilBERT	10	0.4026	87.18
RoBERTa	8	0.3297	90.60

Summary of each model's best epoch, validation loss, and accuracy:

- TF-IDF + LR: Accuracy 81.20% (no epochs/loss).
 - GPT (Prompt-Only): Accuracy 78.60% (no epochs/loss).
 - DistilBERT: Best Epoch=10, Val Loss=0.4026, Val Acc=87.18%.
 - RoBERTa: Best Epoch=8, Val Loss=0.3297, Val Acc=90.60%.
- RoBERTa demonstrates the lowest loss and highest accuracy.

Final Validation Accuracy Comparison

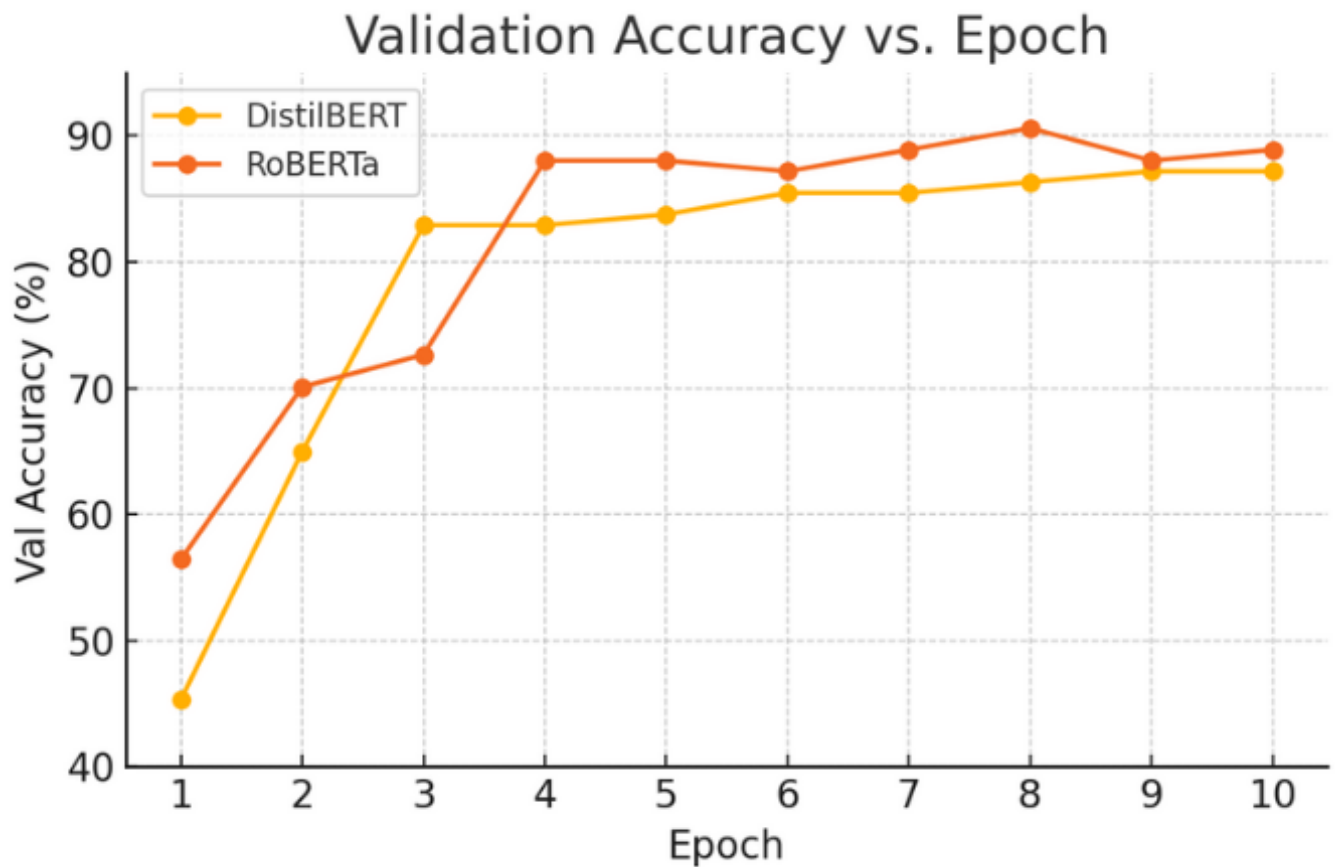


This bar chart compares final validation accuracies across models:

- TF-IDF + LR (Baseline): 81.20%
- GPT (Prompt-Only): 78.60%
- DistilBERT: 87.18%
- RoBERTa: 90.60%

RoBERTa outperforms all other approaches.

Validation Accuracy vs. Epoch



The plot shows validation accuracy vs. epoch for DistilBERT and RoBERTa. RoBERTa (orange) rapidly improves and reaches ~90% at epoch 8, then stabilizes. DistilBERT (yellow) converges more gradually, peaking at 87.18% by epoch 10.