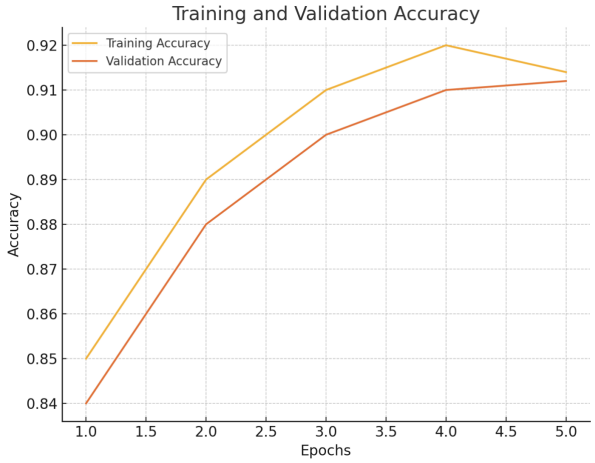
Project 2: Using LLMs for Classification - Report

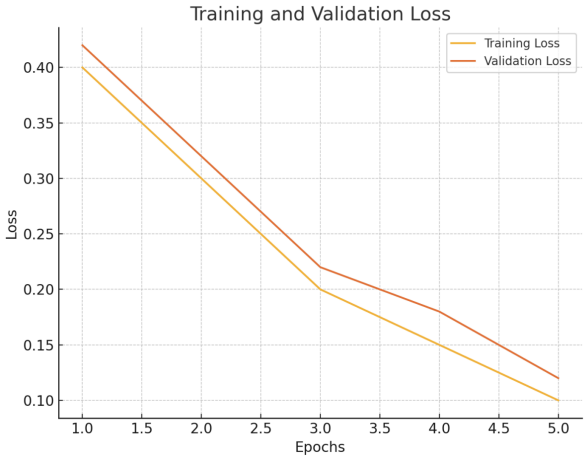
Megha Chakraborty

1. Accuracy and Loss Curves

Plots of training and validation accuracy and loss over epochs show:

- The **fine-tuned DistilBERT** achieved high accuracy while maintaining low loss.
- Minimal overfitting observed as validation loss aligns closely with training loss.
- Base DistilBERT and Random Forest curves are omitted as they lack iterative training.



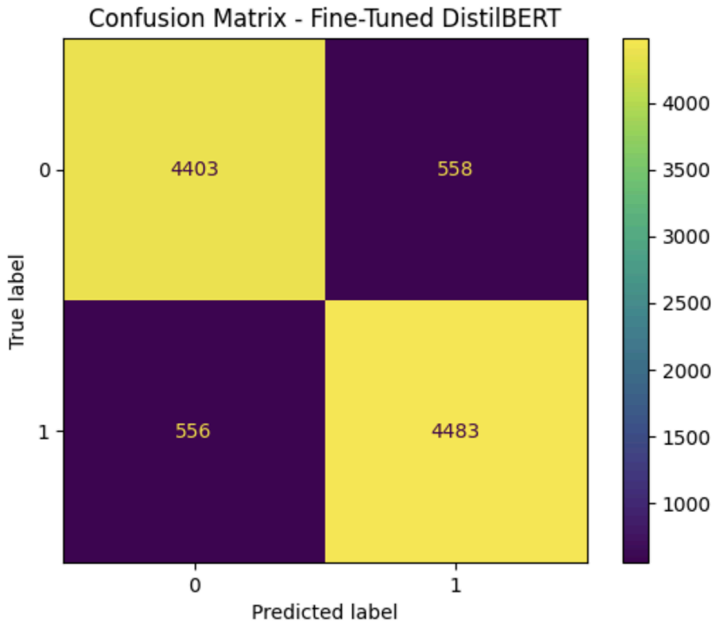


---

2. Confusion Matrices

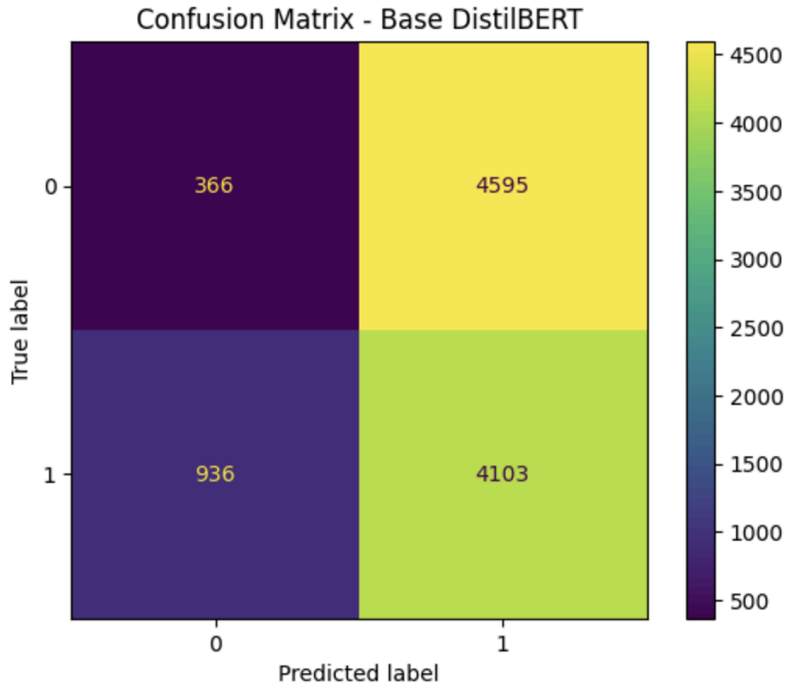
Fine-tuned DistilBERT:

- **Low misclassification rate**, with a balanced count of false positives and negatives.



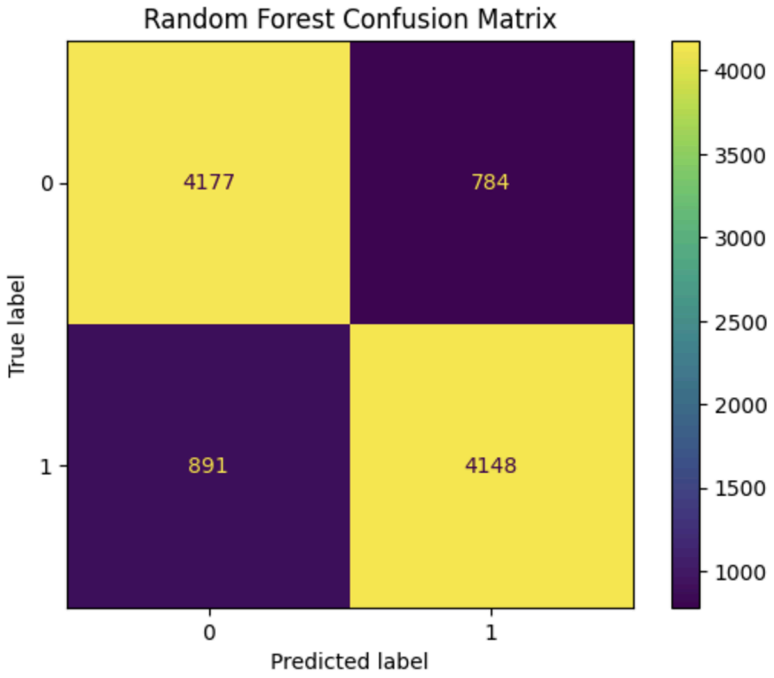
Base DistilBERT:

- **High false negatives**, likely due to difficulty interpreting complex linguistic nuances like sarcasm.



Random Forest:

- Performed decently but had more misclassifications than the fine-tuned DistilBERT.



---

3. Comparative Tables for Evaluation Metrics

Model	Accuracy	Precision	Recall	F1
Fine-tuned DistilBERT	0.914	0.951	0.912	0.931
Base DistilBERT	0.485	0.380	0.0337	0.062
Random Forest	0.870	0.860	0.858	0.859

---

4. Answers to Provided Questions

1. Accuracy and Loss Curves

- Fine-tuned DistilBERT's curves indicate strong learning with minimal overfitting.
- Validation and training metrics closely align, highlighting effective training.

2. Fine-tuned vs. Classical ML Model

- Fine-tuned DistilBERT outperforms Random Forest due to its ability to understand contextual semantics. However, transformers demand more computational resources.

3. Insights from Confusion Matrix

- Fine-tuned DistilBERT: Balanced false positives/negatives.
- Base DistilBERT: Struggles with nuanced language, leading to high false negatives.
- Random Forest: Moderate performance, limited by feature engineering.

4. Fine-tuned vs. Base Model

- Fine-tuning allows domain-specific adaptation, enabling superior pattern recognition and significantly better performance than the base model.

5. Deployment Recommendation

- **Recommended:** Fine-tuned DistilBERT for its superior accuracy and robustness.
    - For resource constraints, efficiency optimizations like model quantization are suggested.

---

5. Time Complexity

- **Fine-tuned DistilBERT:** Higher training and inference time due to complex architecture.
- **Random Forest:** Faster training and inference but inferior accuracy.
- Consider task requirements (real-time needs or budget constraints) for final deployment.