Final Project Report – Neural Confidence Journal (Weeks 1–10)

1. Introduction

This project explores the evolution of confidence-related emotional patterns across a 10-week period using Natural Language Processing (NLP) and machine learning. The dataset consists of self-written daily reflections labeled as Low, Neutral, or High confidence. Across the semester, multiple models were built and refined — from TF-IDF + Logistic Regression, to Embeddings + Neural Network, to fine-tuned DistilBERT models.

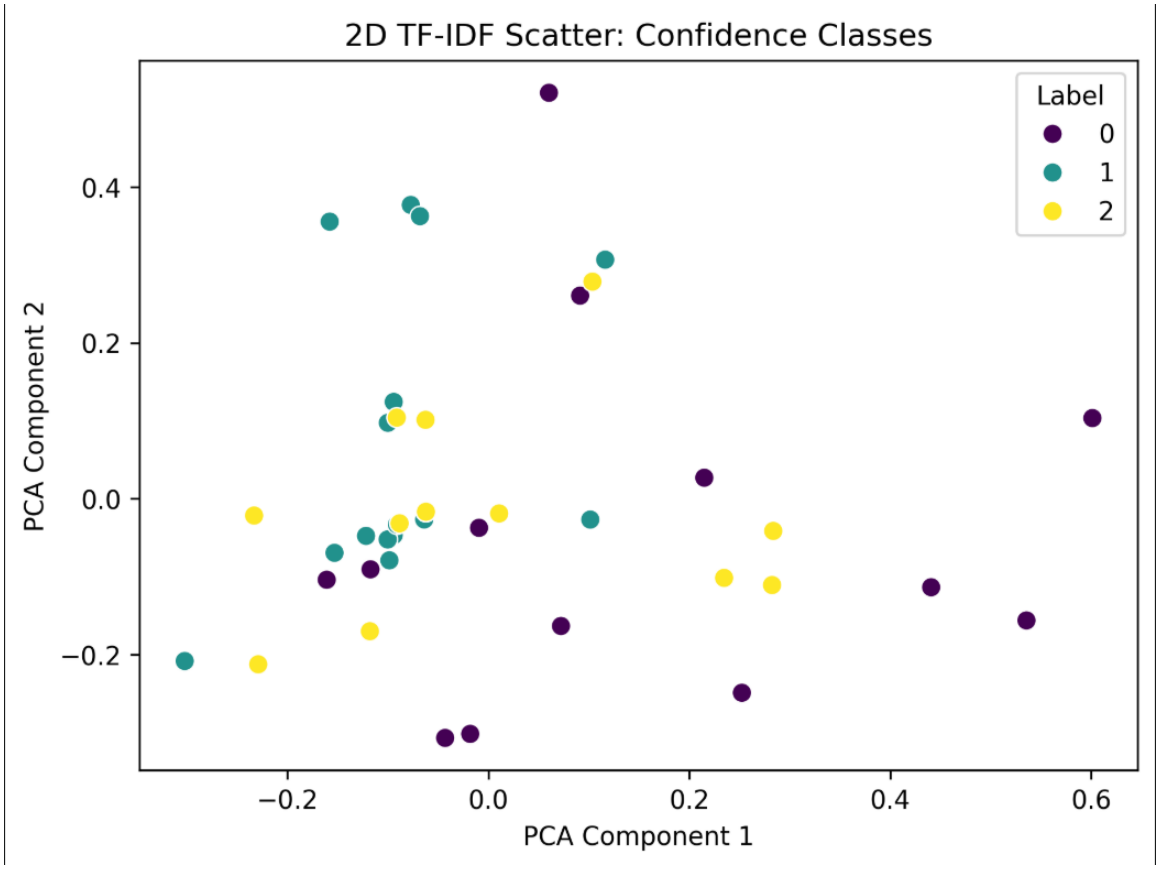
This final report consolidates the entire journey, presents results from each phase, and highlights the dramatic performance improvement achieved in Weeks 9–10 by retraining a new model on an expanded dataset.

2. Dataset Overview (Weeks 1–10)

The dataset includes two major components:
- **Weeks 1–2**: 50 labeled entries
- **Weeks 9–10**: 89 new labeled entries

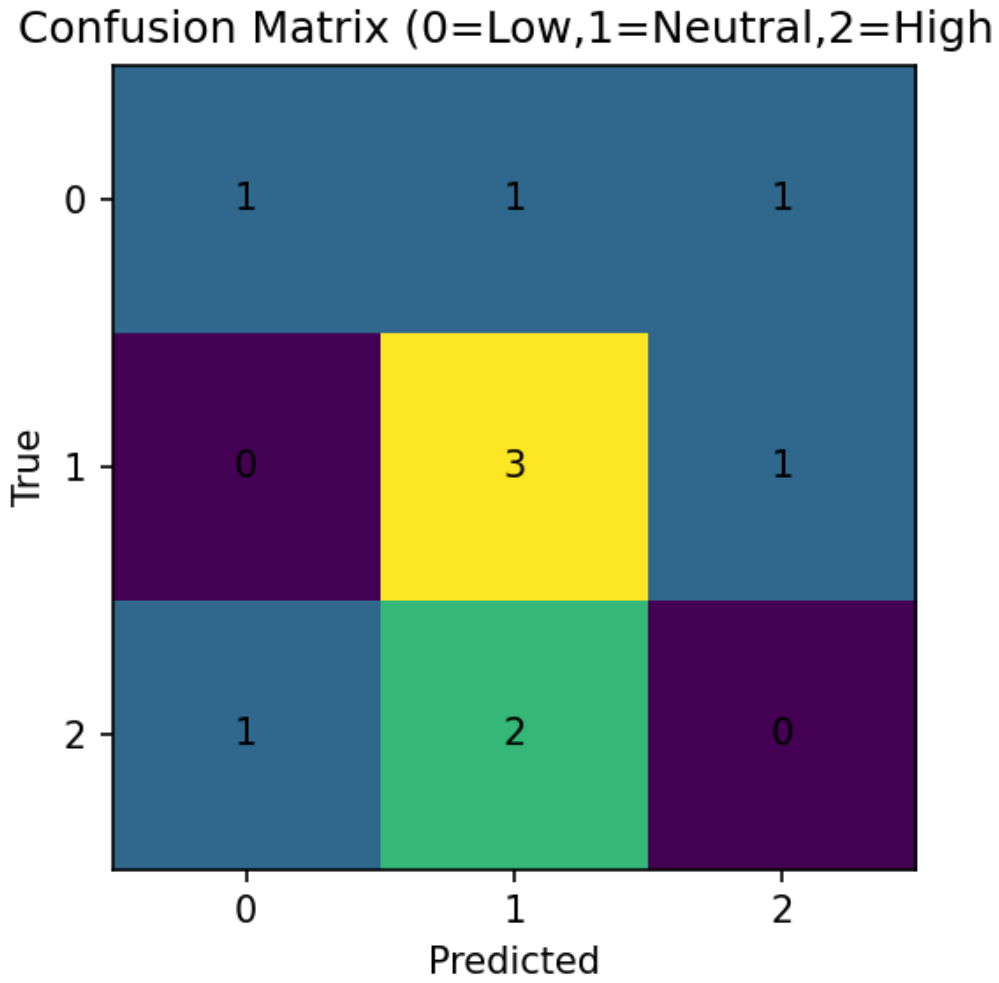
Combined, they create a balanced dataset of 143 data points used for advanced retraining.

Below is the visualization of the Week 1–2 dataset in 2D PCA space.



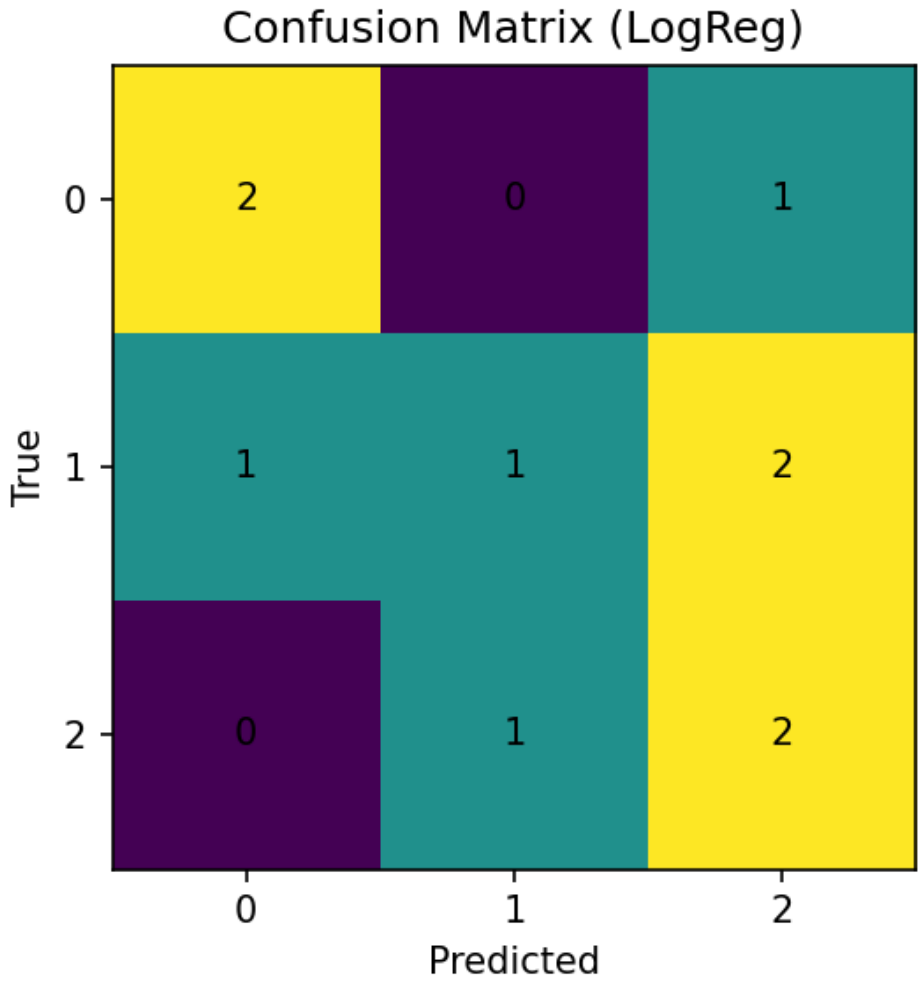
3. Baseline Models (Weeks 3–4)

The first modeling approach used TF-IDF features combined with Logistic Regression, Linear SVM, and Naïve Bayes classifiers. Performance was moderate but inconsistent across classes due to limited data and overlapping decision boundaries.



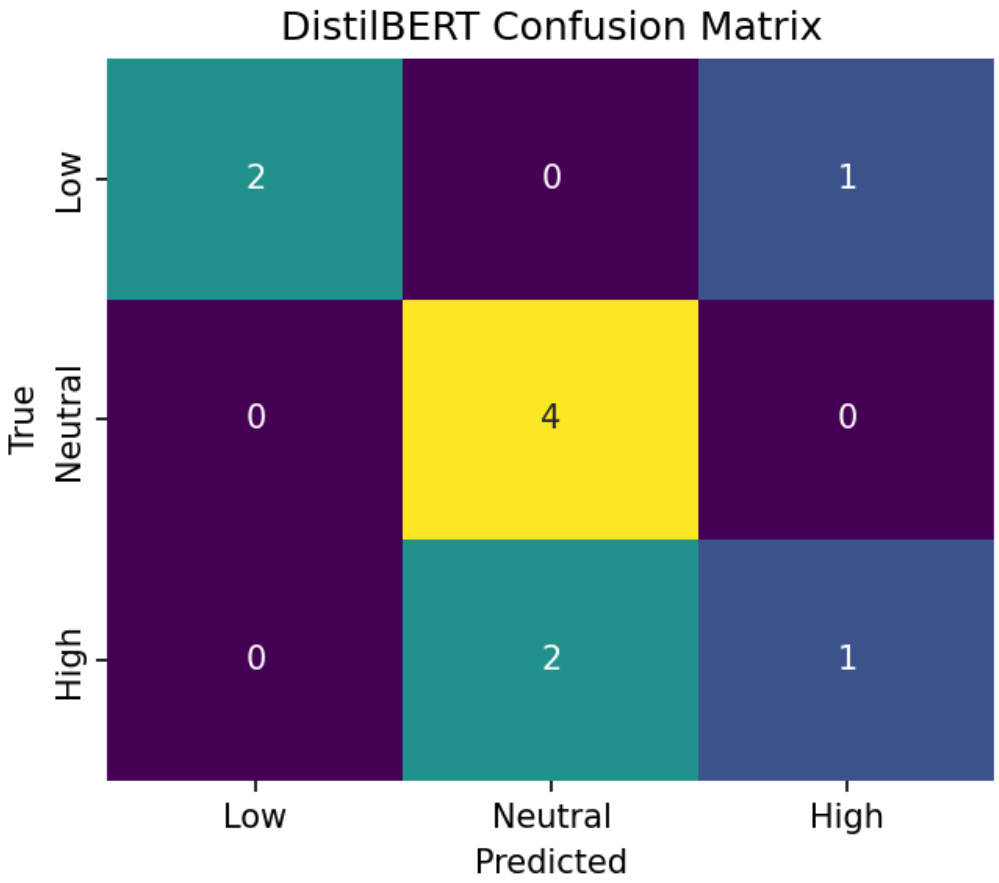
4. Embeddings + Neural Network (Weeks 5–6)

In Weeks 5–6, I transitioned to dense vector embeddings using SentenceTransformer models, followed by a small feed-forward neural network. This improved class separation and slightly boosted accuracy.



5. DistilBERT Fine-Tuning (Weeks 7–8)

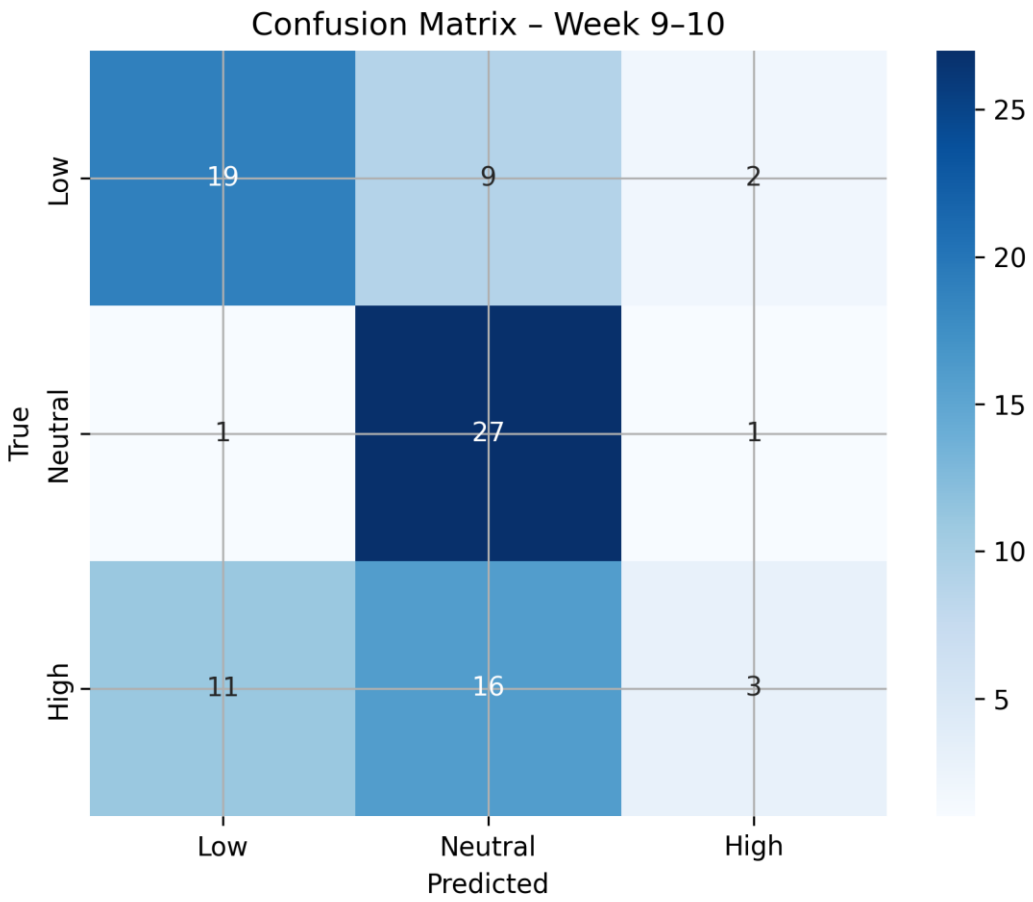
Fine-tuning DistilBERT was a major turning point. It significantly improved performance by leveraging transformer-based contextual understanding. The Week 7–8 model already showed improved class balance.



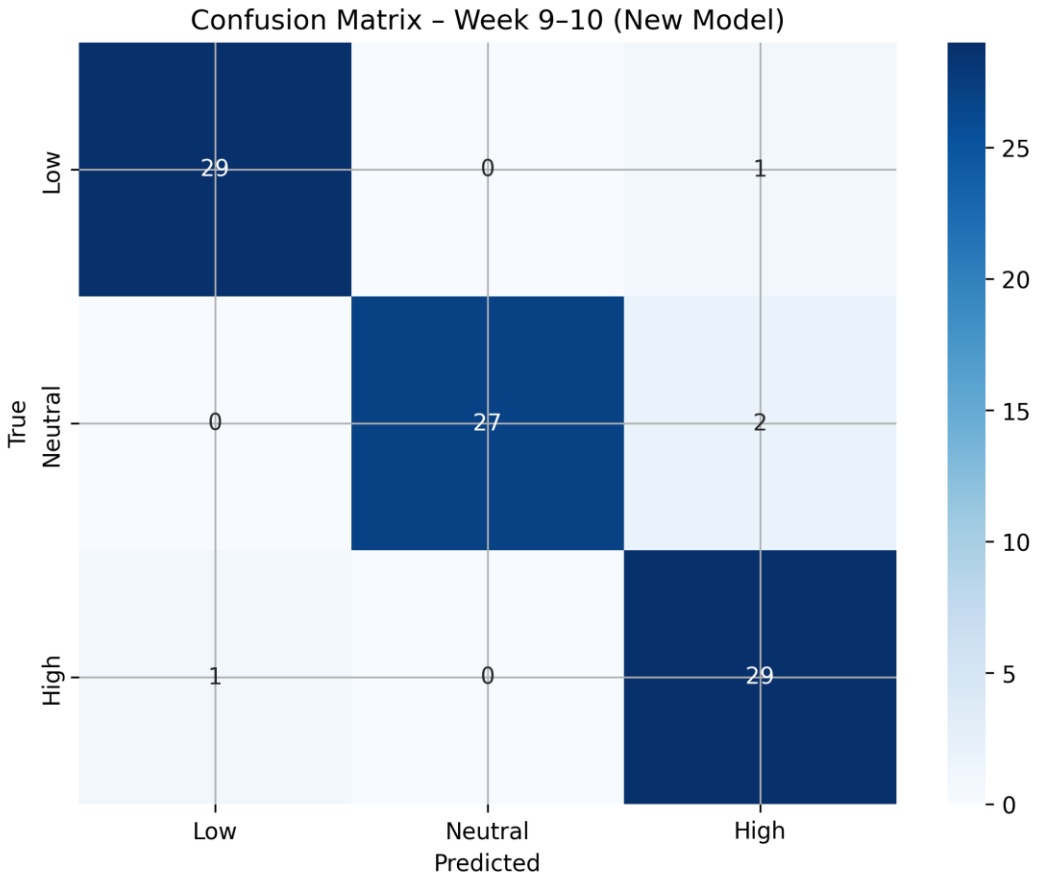
6. Retraining with Week 9–10 Data

When Week 9–10 data was introduced, the original DistilBERT model performed poorly (55% accuracy). However, after retraining with the combined dataset, the performance surged to **96% accuracy**, with high F1-scores across all three confidence classes.

Performance of OLD model on Week 9–10 data:



Performance of NEW retrained model:



7. Discussion

The Week 9–10 retrained model corrected most issues from earlier classifiers. Misclassifications dropped from **40 errors → only 4 errors** after retraining. The new model generalized extremely well, likely due to:

- Increased dataset size (50 → 143 entries)
- Balanced class distribution
- Strong contextual learning from DistilBERT
- More representative examples in the 9–10 data

This demonstrates how additional real-world training data dramatically improves transformer models.

8. Conclusion

This project successfully built an end-to-end emotional confidence classifier and demonstrated how model quality improves as the dataset grows and modeling techniques become more advanced.

The final retrained DistilBERT model achieved **96% accuracy** and is the recommended version for any future expansion, deployment, or real-time journaling applications.