Neural Confidence Journal: Weeks 5–6 Progress Report

Author: Earnest Kyle

---

1. Overview

During Weeks 5–6 (October 7 – October 20), the focus was on improving model performance by transitioning from sparse TF-IDF features to dense **sentence embeddings**. We implemented semantic representations using **MiniLM** from Sentence-Transformers and compared three classifiers:

- **Logistic Regression** (baseline for linear decision boundaries)
- **Linear SVM** (high-margin linear model)
- **Feed-Forward Neural Network (FFNN)** (two dense layers with dropout)

This stage aimed to determine whether embedding-based features could capture subtler emotional cues and confidence expressions in journal entries.

---

2. Methods

Each journal entry was converted into a **384-dimensional MiniLM sentence embedding**. We then used these vectors as input to three classifiers with the following setup:

- **Logistic Regression** – class weight balanced, C = 1.0
- **Linear SVM** – hinge loss, balanced class weights
- **Feed-Forward Neural Network** – 2 dense layers (128 → 64), ReLU activations, dropout = 0.3, softmax output

The dataset was **split 80/20** into train and validation sets using **stratified sampling** to maintain class balance.
All models were evaluated with **accuracy** and **macro F1-score** to account for class imbalance.

---

3. Results

Model	Accuracy	Macro F1	Notes
Logistic Regression	0.50	0.50	Best overall baseline performance
Linear SVM	0.50	0.50	Nearly identical to Logistic Regression
Feed-Forward NN	0.40	—	Underfit due to limited data

The embedding-based traditional models achieved **moderate accuracy (50%)**, showing consistent predictions but limited separation between classes.
The **FFNN reached 40% validation accuracy**, confirming that while embeddings are useful, the dataset size was too small to train a neural classifier effectively.

---

4. Confusion Matrix Insights

The confusion matrix for the Logistic Regression model showed moderate accuracy across all three confidence levels (low, neutral, high).
Class 1 ("neutral") was the most frequent misclassification, overlapping significantly with "high-confidence" entries.
Despite this, overall prediction distribution was more stable compared to TF-IDF models from earlier weeks.

(If you have an image, you can insert the confusion-matrix plot here.)

---

5. Analysis — Why Performance Was Limited

1. **Small Dataset Size:** With fewer than 50 examples, the models cannot generalize robustly.
2. **Semantic Overlap:** Many entries express confidence through similar phrases, making boundary learning difficult.
3. **Static Embeddings:** MiniLM embeddings were not fine-tuned for this domain and therefore capture generic semantic features instead of contextual confidence signals.
4. **Neural Network Underfitting:** The FFNN lacked data to form meaningful non-linear boundaries.
5. **Possible Class Imbalance:** Slight over-representation of certain confidence levels may have biased learning.

---

6. Next Steps (Weeks 7–8 Preview)

The next phase will transition from classical models to transformer-based architectures. Planned actions include:

- **Fine-tune DistilBERT** for direct text classification of confidence labels.
- **Expand and balance the dataset** to at least 300 entries for better representation.
- **Add data augmentation** (e.g., paraphrasing via back-translation).
- **Evaluate with macro F1, recall, and confusion matrices** for consistent comparison.

Expected improvements include a +15–25% increase in accuracy and better class separation.

---

7. Reflection

Weeks 5–6 marked the transition from surface-level text features to semantic understanding through embeddings.
Although overall performance remained modest, the experiments validated that the embedding-based pipeline functions and scales well for future deep learning approaches.
This phase successfully laid the foundation for fine-tuning transformer models in the next stage of the Neural Confidence Journal project.

End of Weeks 5–6 Report
(Next phase: DistilBERT fine-tuning and data augmentation experiments.)