Neural Confidence Journal: Weeks 1–4 Progress Report

Author: Earnest Kyle
Date: October 7, 2025

Project Overview & Goals

The Neural Confidence Journal project explores how language patterns in journal entries reflect varying levels of confidence. Our objective is to build natural language processing (NLP) models capable of classifying journal text into three categories: Low, Neutral, and High confidence. The first four weeks of this project focused on exploratory data analysis (EDA), feature engineering, and baseline model construction.

Weeks 1–2: Exploratory Data Analysis (EDA)

The initial phase concentrated on understanding the dataset and preparing it for modeling. Key steps included: - Examining label distribution (balanced across the three classes) - Analyzing text length and structure - Identifying most common words by confidence level - Creating word clouds to visualize emotional tone - Conducting TF-IDF scatter plots to assess separability This analysis revealed subtle linguistic patterns, such as words like 'worried' or 'felt' correlating with low confidence, while words like 'finished' and 'tomorrow' appeared more frequently in neutral entries.

Weeks 3–4: Baseline Modeling

The next phase focused on building simple baseline models to establish reference performance benchmarks. We implemented two models: Naive Bayes and Logistic Regression, both using TF-IDF vectorization. Results: - Naive Bayes 5-fold macro F1: 0.329 (+/- 0.105) - Logistic Regression (balanced) 5-fold macro F1: 0.368 (+/- 0.164) After tuning Logistic Regression's regularization parameter (C) using a small grid search, the best model achieved: - Accuracy: 0.400 - Macro F1: 0.333 - Weighted F1: 0.360 These results were significantly lower than expected, indicating that the model struggled to differentiate between confidence levels. The confusion matrix confirmed this, showing heavy misclassification, especially for the 'High' class.

Why Did the Model Perform Poorly?

The baseline model’s low accuracy (~40%) and poor F1-scores suggest that simple approaches were insufficient. Several factors likely contributed: - **Small dataset size:** With fewer than 50 samples, the model lacked enough examples to learn robust class boundaries. - **Class overlap:** Vocabulary used in Low, Neutral, and High confidence entries often overlapped, making them harder to distinguish. - **Limited features:** TF-IDF features capture word frequency but miss subtle linguistic cues like syntax, sentiment, or context. - **Imbalanced signal:** Although classes were balanced in count, the semantic differences between them were subtle and not linearly separable.

Next Steps & Recommendations

While baseline performance was modest, it provided critical insights for future improvements. The next steps will include: - Expanding the dataset with more journal entries to improve model generalization. - Experimenting with advanced vectorizers like word embeddings (Word2Vec, GloVe) or contextual models (BERT). - Incorporating sentiment and syntactic features to capture nuanced language patterns. - Trying deep learning architectures and ensemble approaches for improved classification accuracy. Despite the challenges, these foundational experiments established an essential performance baseline and clarified key obstacles that future models must overcome.