Classifying *The Stanford Daily* Headlines Using Naïve Bayes

Melita D'Souza

December 3rd, 2025

1 Introduction

At *The Stanford Daily*, journalists follow a strict rule: you cannot write for both the News section and the Opinions section.

This separation exists to protect journalistic integrity. News writing is expected to be neutral, fact-based, and free of personal judgment. Opinion writing, by contrast, is intentionally argumentative, persuasive, and grounded in the writer's own perspective.

Although this separation is enforced editorially, it raises an interesting qualitative question: is the difference between News and Opinion writing detectable purely from the language itself? In other words, if headlines from each section are written in distinct styles, then statistical patterns in those headlines should be detectable, even to a simple probability model.

As both a writer for *The Daily* and a student studying probability in CS 109, I became interested in a natural question:

> **Are News and Opinion headlines so different that a probability model can reliably tell them apart? And if so, what features make a headline sound argumentative?**

To explore this, I built a Naïve Bayes classifier using only core CS109 concepts, such as Bayes' Rule, conditional independence, relative frequency estimation, maximum likelihood estimation, and log-probabilities. The classifier attempts to categorize *Stanford Daily* headlines as either News or Opinion using nothing more than the words they contain.

2 Dataset

I collected 100 real headlines from *The Stanford Daily*:

- 50 News headlines, representing neutral reporting
- 50 Opinion headlines, representing author-driven argumentation

The headlines span a mix of topics–student government, campus events, politics, research, social issues, administrative changes, and national policy–reflecting the diversity of coverage within *The Daily.*

To evaluate the model fairly and avoid training on the same examples it is tested on, I divided the dataset into two parts:

1. Training Set (80 headlines):
   - 40 News
   - 40 Opinion

   These headlines were used to estimate word frequencies and compute the conditional probabilities required by the Naïve Bayes classier.
2. Test Set (20 headlines):
   - 10 News
   - 10 Opinion

   These headlines were held out entirely from training and used only to measure the model's final accuracy.

3 Method

We want to compute $P(\text{News} \mid H)$ and $P(\text{Opinion} \mid H)$ for a new headline H.

Bayes' Rule states $P(C \mid H) \propto P(H \mid C)\,P(C)$ with with $C \in \{\text{News}, \text{Opinion}\}$. Since headlines contain only a few words, we model the probability of a headline as:
$P(H \mid C) = \prod_{w \in H} P(w \mid C)$

This naïve independence assumption is exactly the one discussed in CS 109 during text-classification examples.

All probabilities were estimated using word frequencies from the training set:
$P(w \mid C) = \frac{\text{count}(w \text{ in class } C)}{\text{total word count in class } C}$

Priors were estimated as: $P(\text{News}) = P(\text{Opinion}) = 0.5$
To avoid underflow, I summed logs: $\log P(C \mid H) = \log P(C) + \sum_{w \in H} \log P(w \mid C)$
Finally, the model predicts the class with the larger log-probability.

The code can be found here: `https://github.com/melitasdsouza/stanford-daily-naive-bayes/blob/main/classifier.py`

4 Results

To evaluate how well the Naïve Bayes classifier generalizes to unseen headlines, I tested it on a held-out set of 20 Stanford Daily headlines (10 News and 10 Opinion) that were not used during training. The classifier correctly labeled 17 out of 20 headlines, achieving an overall accuracy of 85%.

The breakdown of predictions is shown in Table 1.

	Predicted News	**Predicted Opinion**
Actual News	9	1
Actual Opinion	2	8

Table 1: Confusion matrix for Naïve Bayes classifier on Stanford Daily headlines.

The classifier performed slightly better on News headlines, correctly identifying 9 out of 10. The single News misclassification occurred when the headline included vocabulary more commonly associated with Opinion writing (e.g., terms linked to activism or social movements).

Performance on Opinion headlines was slightly lower, with 8 out of 10 correctly classified. The two Opinion headlines that were mislabeled tended to include more institutional or policy-oriented language, which made them appear more similar to News headlines.

Overall, the classifier's performance suggests that News and Opinion headlines exhibit distinct linguistic patterns that are measurable through word frequencies alone, even without context, smoothing, or more advanced language modeling. These quantitative differences support the hypothesis that the stylistic boundary between News and Opinion is strong enough that a purely probabilistic model can identify it with high accuracy.

5 Error Analysis

Although the classifier performed well overall, its mistakes reveal meaningful linguistic patterns. To better understand these errors, I examined which words were most distinctive of each section in the training data. Figures 2 and 1 show the top words that appeared frequently in one section but rarely (or never) in the other.
On the Opinion side (Figure 1), words such as why, not, we, and how occur frequently. These are rhetorical or argumentative markers that commonly indicate a persuasive or value-driven headline.

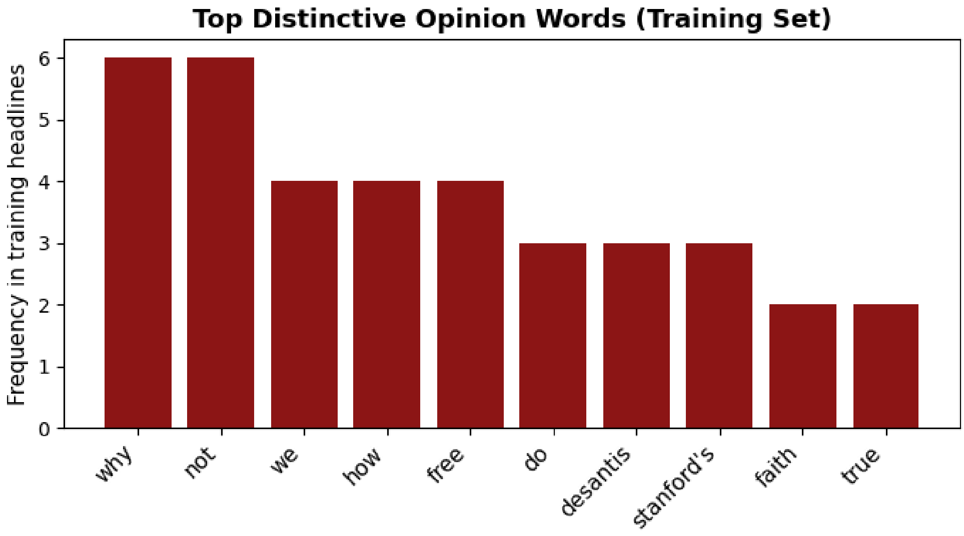

Figure 1: Top distinctive Opinion words in the training set. These words appear frequently in Opinion headlines and rarely in News headlines, reflecting the rhetorical and evaluative style typical of Opinion writing.

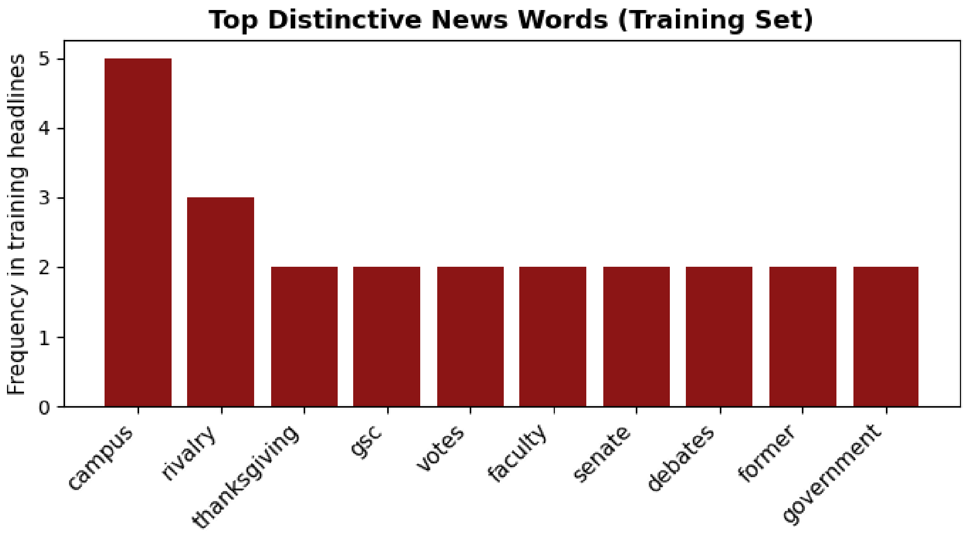

Figure 2: Top distinctive News words in the training set. These words reflect institutional, event-oriented, and descriptive language common in News reporting.

Their presence reflects the personal, evaluative tone characteristic of Opinion writing.

On the News side (Figure 2), words such as campus, rivalry, faculty, votes, and senate appear prominently. These terms refer to institutions, events, and groups—hallmarks of fact-based reporting. News headlines tend to focus on concrete entities rather than rhetorical structure.

These patterns help explain the classifier's errors. For example, the misclassified Opinion headlines often contained institutional vocabulary (e.g., policy or governance terms), causing them to resemble News headlines. Conversely, the misclassified News headline included words associated with activism or reflection, making it appear stylistically closer to Opinion.

Together, the distinctive-word distributions show that a small number of discriminative words drive much of the classifier's predictive power, highlighting the stylistic boundary between the two sections.