

Advanced Natural Language Processing

Assignment 1: Twitter hate speech detection using Naive Bayes and Logistic Regression

Lucia Welther
Matrikelnummer: XXXXXXXX

November 16, 2025

1 Smoothing Parameter Analysis

Figure 1 shows the relationship between the smoothing parameter k and model performance.

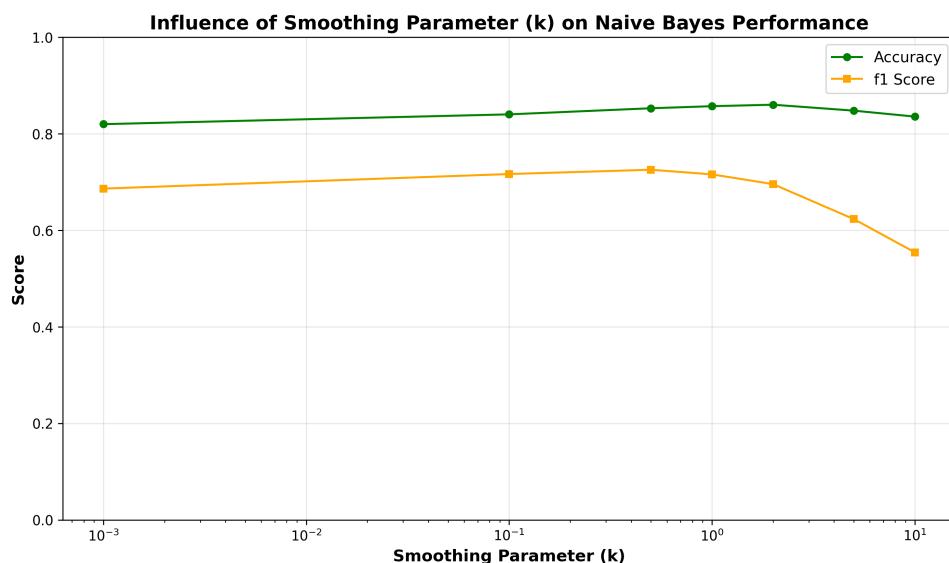


Figure 1: Accuracy and F1-score as functions of smoothing parameter k

Observations:

- When k is very small (e.g., 0.001), the model shows [describe performance - e.g., "lower performance due to overfitting to the training data"].
- As k increases to around [X], performance [increases/stabilizes] because...
- When k becomes very large (e.g., 10), performance [decreases/remains stable] because...

2 Feature Engineering

2.1 Feature Set 1: Stopword Removal

Results:

- Accuracy: [your value]
- F1-score: [your value]

2.2 Feature Set 2: Bigrams

Results:

- Accuracy: [your value]
- F1-score: [your value]

2.3 Comparison

Table 1 summarizes the performance of different feature sets.

Table 1: Performance comparison of feature engineering approaches

Feature Set	Accuracy	F1-Score
Baseline (Bag-of-Words)	X.XXXX	X.XXXX
Binary BoW (Feature 1)	X.XXXX	X.XXXX
Bigrams (Feature 2)	X.XXXX	X.XXXX

3 Model Evaluation Results

3.1 Naive Bayes Classifier

The Naive Bayes classifier with optimal smoothing parameter $k = [your value]$ achieved:

- **Accuracy:** [your value]
- **F1-Score:** [your value]

3.2 Logistic Regression Classifier

The Logistic Regression classifier with learning rate $\eta = 0.01$, 10 epochs, and L2 regularization parameter $C = 0.1$ achieved:

- **Accuracy:** [your value]
- **F1-Score:** [your value]

3.3 Overall Comparison

Table 2 compares the two classifiers.

Table 2: Performance comparison of Naive Bayes vs. Logistic Regression

Model	Accuracy	F1-Score
Naive Bayes	X.XXXX	X.XXXX
Logistic Regression	X.XXXX	X.XXXX

4 Bonus: Discussion Questions

4.1 Effect of Smoothing Parameter

[Answer Question 1: What is the effect of the smoothing parameter k in Naive Bayes for the calculation of word probabilities? What happens to the probabilities of the words when k is very large compared to their observed frequencies in the training data?]

4.2 Word Order Shuffling

[Answer Question 2: If we shuffle the order of the words within each document during training, will this affect the results of the model on the test set? Why or why not?]

4.3 Purpose of Softmax Function

[Answer Question 3: Why do we need the softmax function if the model already produces a score for each class?]

4.4 Decision Threshold Tuning

[Answer Question 4: In this binary classification setup we select the predicted class by taking the arg max over the class scores. In this case is equivalent to choosing the class whose predicted probability is greater than 0.5. If we wanted to fine-tune the decision threshold to improve the evaluation results: how many times (if at all) do we need to train the model to find the best threshold value?]

4.5 GitHub Repository

The complete implementation is available at: <https://github.com/yourusername/nlp-assignment1>
Last commit hash: d

5 Conclusion