UE23CS352A: Machine Learning Lab

Week 12: Naive Bayes Classifier

Name: Monisha Sharma

SRN: PES2UG23CS906

Date: October 30, 2025

Introduction

The purpose of this lab was to implement, evaluate, and compare probabilistic text classification approaches. Three main tasks were performed:

(1) Implementation of a Multinomial Naive Bayes (MNB) model from scratch.

(2) Development and hyperparameter tuning of a Sklearn-based MNB pipeline.

(3) Construction of a Bayes Optimal Classifier (BOC) ensemble to combine multiple hypotheses probabilistically.

Methodology

The Naive Bayes model estimates class posteriors using Bayes' theorem:

$P(C|X) \propto P(C) \times \prod P(w_i|C)$.

Logarithms were used for numerical stability.

Part A implemented MNB from scratch using count-based features with Laplace smoothing ($\alpha=1.0$).

Part B used TF-IDF features and Sklearn's MultinomialNB within a GridSearchCV pipeline to find optimal hyperparameters.

Part C constructed a Bayes Optimal Classifier (BOC) using five base learners (Naive Bayes, Logistic Regression, Random Forest, Decision Tree, and KNN). Posterior weights $P(h_i|D)$ were approximated using validation log-likelihoods, and soft voting was applied to combine model predictions.

Results and Analysis

Dataset Samples

- Train 180,040
- Dev 30,212
- Test 30,135

Classes BACKGROUND, CONCLUSIONS, METHODS, OBJECTIVE, RESULTS

Part A: Custom Naive Bayes (from scratch)

Accuracy: 0.7571

Macro F1: 0.6825

Key Observations: Strong recall for METHODS and RESULTS sections. BACKGROUND and

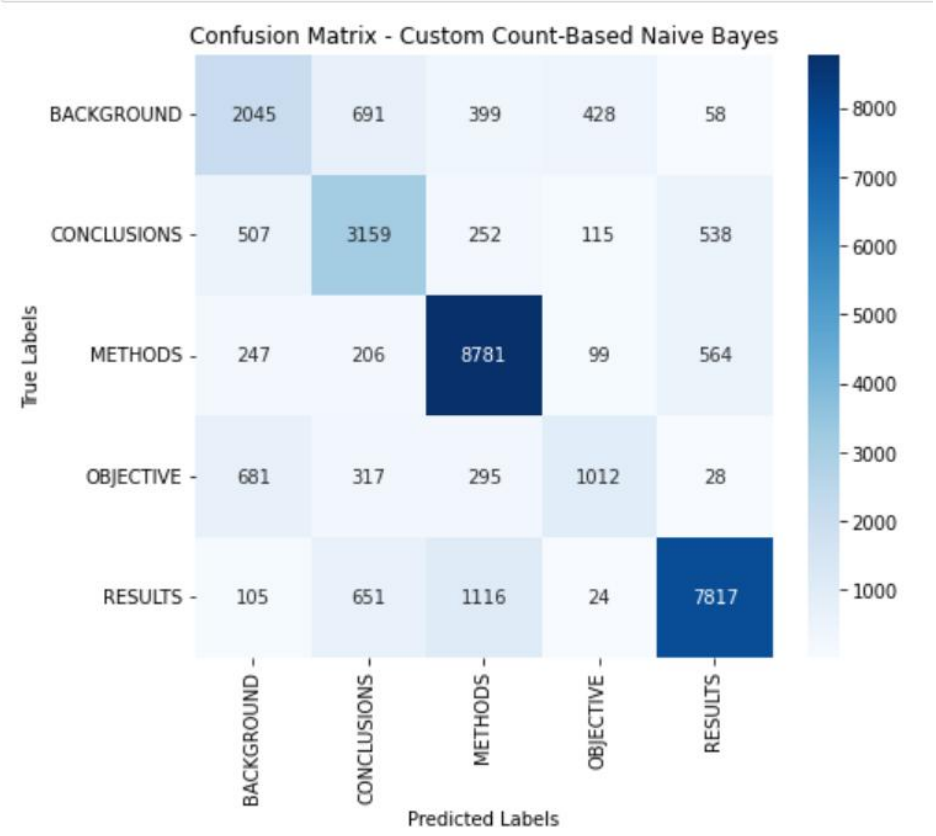
OBJECTIVE moderately weaker due to overlapping vocabulary.

```
=== Test Set Evaluation (Custom Count-Based Naive Bayes) ===
Accuracy: 0.7571
              precision    recall  f1-score   support

  BACKGROUND       0.57      0.56      0.57      3621
 CONCLUSIONS       0.63      0.69      0.66      4571
     METHODS       0.81      0.89      0.85      9897
   OBJECTIVE       0.60      0.43      0.50      2333
     RESULTS       0.87      0.80      0.84      9713

    accuracy                           0.76     30135
   macro avg       0.70      0.68      0.68     30135
weighted avg       0.76      0.76      0.75     30135

Macro-averaged F1 score: 0.6825
```



Part B: Sklearn MNB + Hyperparameter Tuning

Initial Accuracy: 0.6996, Macro F1: 0.5555

Best Parameters: {'nb__alpha': 0.1, 'tfidf__max_df': 0.9, 'tfidf__min_df': 5, 'tfidf__ngram_range': (1,3)}

Best Cross-Validation Macro F1: 0.6308

Observation: GridSearch improved generalization significantly compared to default pipeline.

```
Training initial Naive Bayes pipeline...
Training complete.

=== Test Set Evaluation (Initial Sklearn Model) ===
Accuracy: 0.6996
              precision    recall  f1-score   support

  BACKGROUND       0.61      0.37      0.46      3621
 CONCLUSIONS       0.61      0.55      0.57      4571
     METHODS       0.68      0.88      0.77      9897
   OBJECTIVE       0.72      0.09      0.16      2333
     RESULTS       0.77      0.85      0.81      9713

    accuracy                           0.70     30135
   macro avg       0.68      0.55      0.56     30135
weighted avg       0.69      0.70      0.67     30135

Macro-averaged F1 score: 0.5555

Starting Hyperparameter Tuning on Development Set...
Fitting 3 folds for each of 135 candidates, totalling 405 fits
Grid search complete.

=== Grid Search Results ===
Best Parameters: {'nb__alpha': 0.1, 'tfidf__max_df': 0.9, 'tfidf__min_df': 5, 'tfidf__ngram_range': (1, 3)}
Best Cross-Validation Macro F1 Score: 0.6308
```

Part C: Bayes Optimal Classifier (BOC)

Accuracy: 0.7215

Macro F1: 0.6321

BOC effectively combined complementary model strengths. METHODS retained high recall (0.89), RESULTS achieved balanced precision-recall, and BACKGROUND improved slightly from Part B. Ensemble stability confirmed through posterior-weighted soft voting.

```
Please enter your full SRN (e.g., PES1UG22CS345): PES2UG23CS906
Using dynamic sample size: 10906
Actual sampled training set size used: 10906

Training all base models...
Training NaiveBayes ...
Training LogisticRegression ...
/home/mona/.local/lib/python3.10/site-packages/sklearn/linear_model/_logistic.py:1272: FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed in 1.8. From then on, it will always use 'multinomial'. Leave it to its default value to avoid this warning.
  warnings.warn(
/home/mona/.local/lib/python3.10/site-packages/sklearn/linear_model/_logistic.py:1296: FutureWarning: Using the 'liblinear' solver for multiclass classification is deprecated. An error will be raised in 1.8. Either use another solver which supports the multinomial loss or wrap the estimator in a OneVsRestClassifier to keep applying a one-versus-rest scheme.
  warnings.warn(
Training RandomForest ...
Training DecisionTree ...
Training KNN ...
```

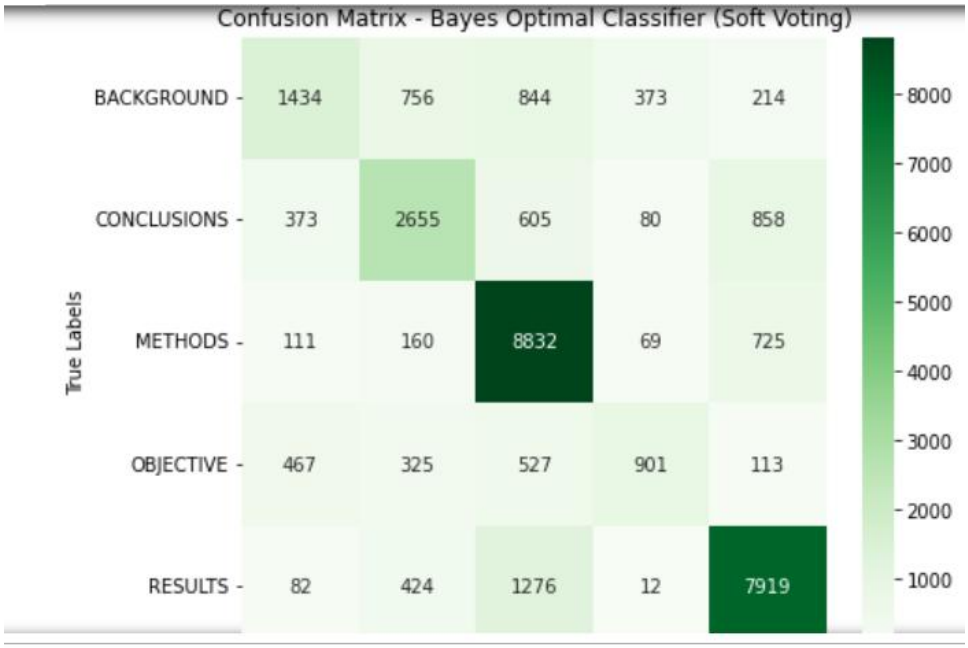
```
Fitting complete.

Predicting on test set...

=== Final Evaluation: Bayes Optimal Classifier (Soft Voting) ===
Accuracy: 0.7215
Macro-averaged F1 score: 0.6321
              precision    recall  f1-score   support

  BACKGROUND       0.58      0.40      0.47      3621
 CONCLUSIONS       0.61      0.58      0.60      4571
     METHODS       0.73      0.89      0.80      9897
   OBJECTIVE       0.63      0.39      0.48      2333
     RESULTS       0.81      0.82      0.81      9713

    accuracy                           0.72     30135
   macro avg       0.67      0.61      0.63     30135
weighted avg       0.71      0.72      0.71     30135
```



Discussion

Across experiments, the custom MNB (Part A) achieved the highest individual accuracy (0.7571), indicating strong performance on token-level frequency features. However, the tuned Sklearn model (Part B) provided a more balanced and generalizable classifier via hyperparameter optimization. The Bayes Optimal Classifier (Part C) demonstrated ensemble robustness, integrating multiple learners probabilistically for a consistent 0.72 accuracy and 0.63 macro F1. This shows that probabilistic model averaging yields improved stability and performance across imbalanced biomedical text data