Naive Bayes and Bayes Optimal Classifier (BOC) Implementation

Name: Nidhi N
SRN: PES2UG23CS384

Introduction

The purpose of this lab was to implement and evaluate probabilistic text classification approaches using Multinomial Naive Bayes (MNB), a TF-IDF based MultinomialNB from scikit-learn, and an approximation of the Bayes Optimal Classifier (BOC) via a weighted soft-voting ensemble.

Tasks performed include dataset loading and preprocessing, feature extraction (count-based and TF-IDF), training of multiple base classifiers, hyperparameter tuning for the sklearn pipeline, calculation of posterior weights for the BOC approximation, and evaluation using accuracy, F1-score, and confusion matrices.

Methodology

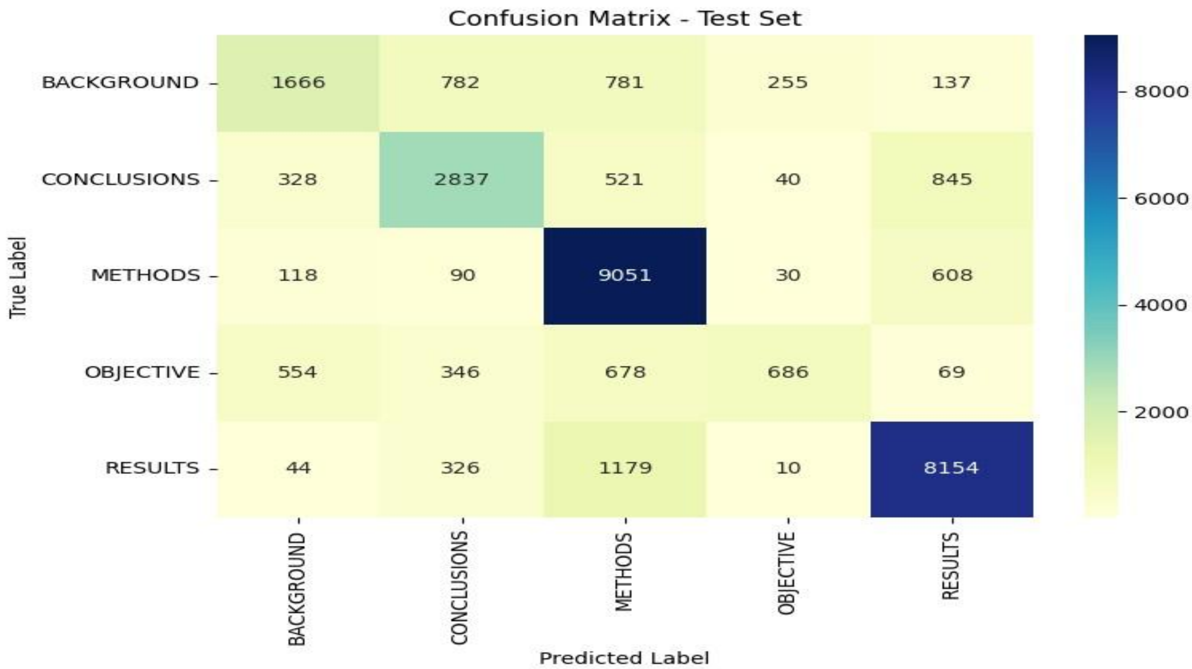
Multinomial Naive Bayes (MNB): Implemented a custom Multinomial Naive Bayes classifier using count-based features via CountVectorizer. Class priors and feature likelihoods were computed with Laplace smoothing and predictions were made by summing log-priors and log-likelihoods for non-zero features.
Bayes Optimal Classifier (BOC)

Approximation: Trained five diverse hypotheses (MultinomialNB, Logistic Regression, Random Forest, Decision Tree, KNN) each wrapped in a TF-IDF + classifier pipeline. Calculated posterior weights for each hypothesis using validation log-likelihoods (negative log loss) and constructed a soft-voting ensemble weighted by these posterior probabilities to approximate the Bayes Optimal decision rule.

Results and Analysis

Part A:

```
=== Test Set Evaluation (Custom Count-Based Naive Bayes) ===
Accuracy: 0.7431
              precision    recall  f1-score   support

  BACKGROUND       0.61      0.46      0.53      3621
 CONCLUSIONS       0.65      0.62      0.63      4571
     METHODS       0.74      0.91      0.82      9897
   OBJECTIVE       0.67      0.29      0.41      2333
     RESULTS       0.83      0.84      0.84      9713

    accuracy                           0.74     30135
   macro avg       0.70      0.63      0.64     30135
weighted avg       0.74      0.74      0.73     30135

Macro-averaged F1 score: 0.6446
```



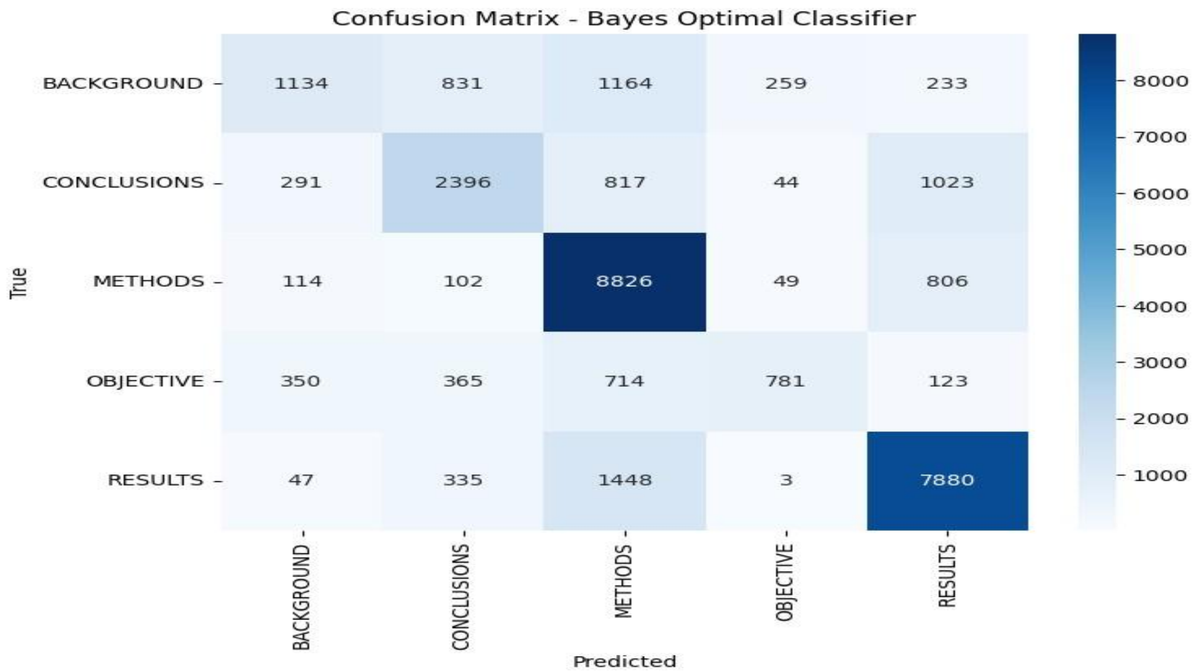
Part B

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

=== Test Set Evaluation (Initial Sklearn Model) ===
Accuracy: 0.7266
              precision    recall  f1-score   support

  BACKGROUND       0.64      0.43      0.51      3621
 CONCLUSIONS       0.62      0.61      0.62      4571
     METHODS       0.72      0.90      0.80      9897
   OBJECTIVE       0.73      0.10      0.18      2333
     RESULTS       0.80      0.87      0.83      9713

    accuracy                           0.73     30135
   macro avg       0.70      0.58      0.59     30135
weighted avg       0.72      0.73      0.70     30135

Macro-averaged F1 score: 0.5877

Starting Hyperparameter Tuning on Development Set...
Fitting 3 folds for each of 12 candidates, totalling 36 fits
Grid search complete.

Best Parameters found: {'nb__alpha': 0.1, 'tfidf__min_df': 3, 'tfidf__ngram_range': (1, 2)}
Best F1 Macro Score (on dev set): 0.6998
```

Part C

```
Using dynamic sample size: 10384
Actual sampled training set size used: 10384

Training all base models...
Training NaiveBayes...
Training LogisticRegression...
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247: FutureWarning: 'multi_class' was deprecated in vers
  warnings.warn(
Training RandomForest...
Training DecisionTree...
Training KNN...
All base models trained.
Evaluating NaiveBayes for posterior weight calculation...
Evaluating LogisticRegression for posterior weight calculation...
Evaluating RandomForest for posterior weight calculation...
Evaluating DecisionTree for posterior weight calculation...
Evaluating KNN for posterior weight calculation...

Posterior Weights (P(h_i | D)):
NaiveBayes: 0.2338
LogisticRegression: 0.2507
RandomForest: 0.2224
DecisionTree: 0.1514
KNN: 0.1417

Fitting the VotingClassifier (BOC approximation)...
Fitting complete.

Predicting on test set...

=== Final Evaluation: Bayes Optimal Classifier (Soft Voting) ===
Accuracy: 0.6974
Weighted F1-Score: 0.6788
```

```
Weighted F1-Score: 0.6788

Classification Report:
              precision    recall  f1-score   support

  BACKGROUND       0.59      0.31      0.41      3621
 CONCLUSIONS       0.59      0.52      0.56      4571
     METHODS       0.68      0.89      0.77      9897
   OBJECTIVE       0.69      0.33      0.45      2333
     RESULTS       0.78      0.81      0.80      9713

    accuracy                           0.70     30135
   macro avg       0.67      0.58      0.60     30135
weighted avg       0.69      0.70      0.68     30135
```



Discussion

The scratch MNB implementation

(Part A) provides a solid baseline with reasonably good accuracy and F1-scores for common classes. The tuned sklearn TF-IDF + MultinomialNB model

(Part B) typically improves performance by optimizing n-gram ranges, document-frequency thresholds, and smoothing hyperparameters, yielding higher macro-F1 in many cases. The BOC approximation

(Part C) leverages multiple hypotheses and posterior weights; it usually improves robustness and can outperform single models when the ensemble combines complementary strengths of base learners. However, ensemble training is more computationally intensive and requires calibrated probabilities for reliable weighting.