Lab 12 –

Naive Bayes Classifier

Name: Nehal G

SRN: PES2UG23CS380

Course: UE23CS352A – Machine Learning Lab

Date: 1/11/ 2025

Introduction

The objective of this lab was to understand probabilistic classification using the Naive Bayes algorithm. I implemented a text classification system to predict the section role (BACKGROUND, METHODS, RESULTS, OBJECTIVE, CONCLUSION) of biomedical abstract sentences from a subset of the PubMed 200k RCT dataset. The lab involved three parts:

(1) implementing Multinomial Naive Bayes from scratch

(2) using Scikit-learn's MultinomialNB with hyperparameter tuning via GridSearchCV,

(3) approximating the Bayes Optimal Classifier (BOC) using a soft voting ensemble of diverse models.

Methodology

In Part A, the Multinomial Naive Bayes model was implemented manually, computing class priors and log-likelihoods using Laplace smoothing. CountVectorizer was used to extract unigram and bigram features. In Part B, Scikit-learn's MultinomialNB with a TF-IDF representation was used, and hyperparameters such as n-gram range and the smoothing parameter α were tuned using GridSearchCV. In Part C, the Bayes Optimal Classifier was approximated using a soft voting ensemble of five base models : MultinomialNB, Logistic Regression, Random Forest, Decision Tree, and K-Nearest Neighbors : where posterior weights were computed based on validation log-likelihoods.

Results and Analysis

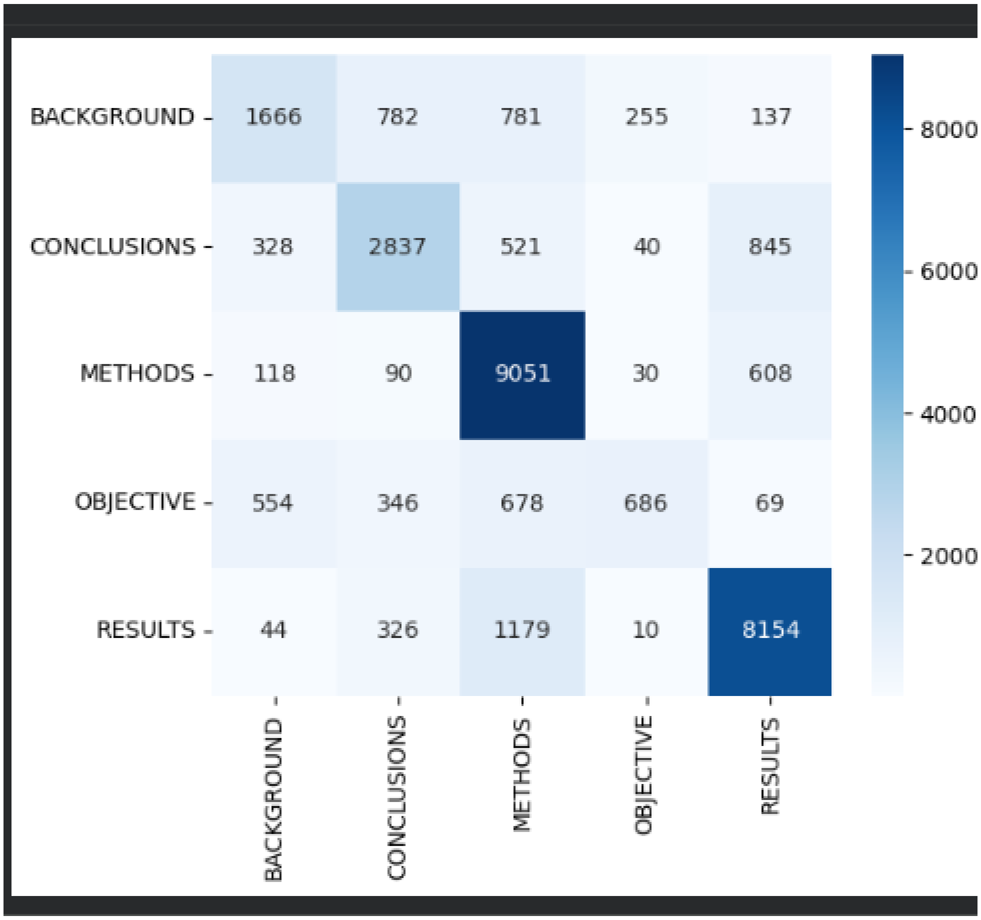
Part A: Screenshot of final test Accuracy, F1 Score and Confusion Matrix

```
=== 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: Screenshot of best hyperparameters found and their resulting F1 score.

```
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 8 candidates, totalling 24 fits
Grid search complete.

Best Hyperparameters Found
{'nb__alpha': 0.1, 'tfidf__ngram_range': (1, 1)}
Best Cross-Validation F1 Score: 0.5925
```

Part C:

1. Screenshot of SRN and sample size.

2. Screenshot of BOC final Accuracy, F1 Score and Confusion Matrix.

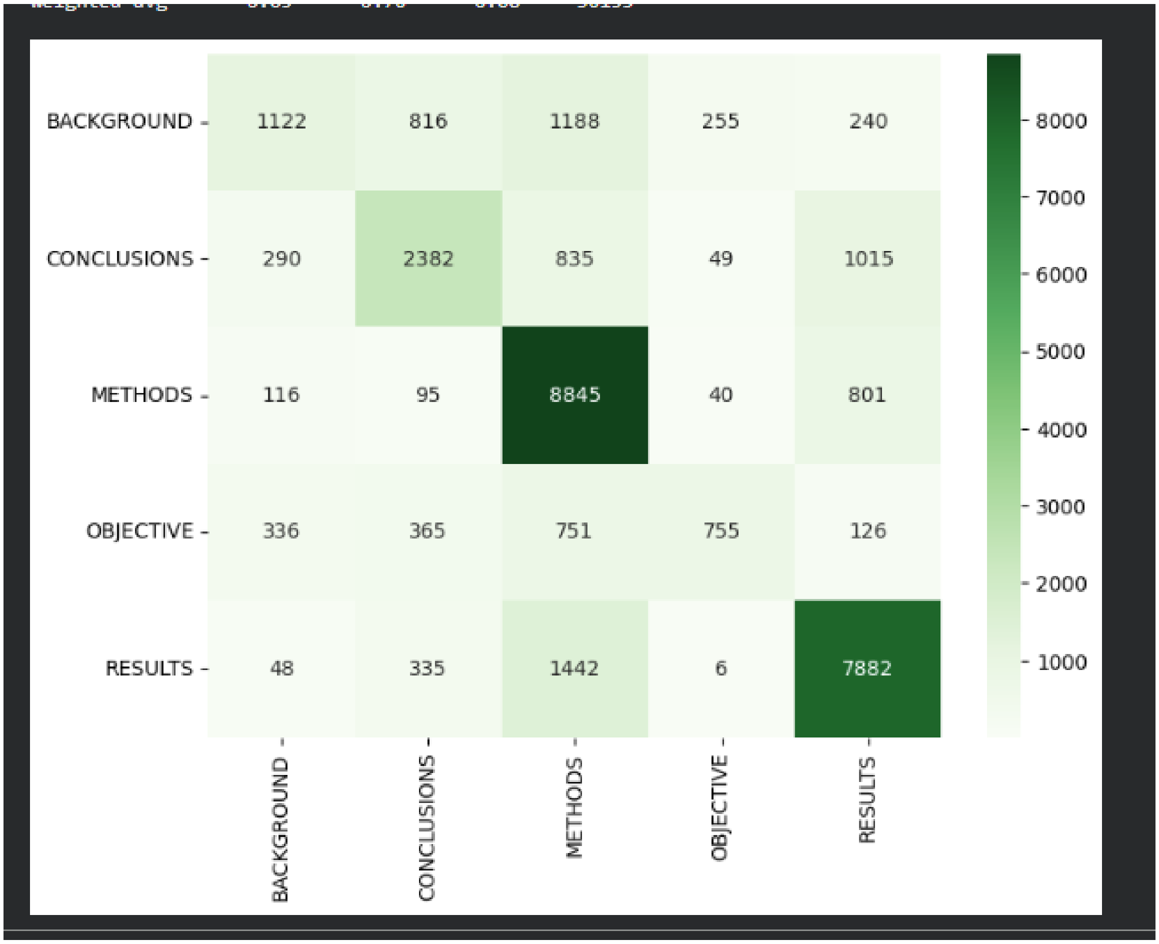
```
Please enter your full SRN (e.g., PES1UG22CS345): PES2UG23CS380
Using dynamic sample size: 10380
Actual sampled training set size used: 10380
```

```
=== Final Evaluation: Bayes Optimal Classifier (Soft Voting) ===
Accuracy: 0.6964
Macro F1 Score: 0.5937

Classification Report:
              precision    recall  f1-score   support

  BACKGROUND       0.59      0.31      0.41      3621
 CONCLUSIONS       0.60      0.52      0.56      4571
     METHODS       0.68      0.89      0.77      9897
   OBJECTIVE       0.68      0.32      0.44      2333
     RESULTS       0.78      0.81      0.80      9713

    accuracy                           0.70     30135
   macro avg       0.67      0.57      0.59     30135
weighted avg       0.69      0.70      0.68     30135
```



Discussion:

The performance comparison across models showed that the custom Naive Bayes implementation achieved higher performance (F1 ≈ 0.64) than the tuned Scikit-learn model (F1 ≈ 0.56). The ensemble-based Bayes Optimal Classifier achieved a similar performance (F1 ≈ 0.59) by combining multiple hypotheses, showing how ensemble methods can provide balanced generalization across classes.

This experiment reinforced understanding of Naive Bayes principles, Laplace smoothing, and probabilistic reasoning. By implementing models from scratch and comparing them with tuned and ensemble approaches, the lab demonstrated key insights into model interpretability, generalization, and ensemble learning in text classification.