ML LAB REPORT – 9

NAME: Mekala Babu Abhinav

SRN: PES2UG23CS338

SECTION: F

DATE: 30/10/2025

Introduction:

The purpose of this lab was to implement and evaluate probabilistic text classification using the Multinomial Naive Bayes (MNB) algorithm on biomedical abstract sentences. The tasks included implementing the MNB classifier from scratch (Part A), utilizing scikit-learn's Naive Bayes with hyperparameter tuning (Part B), and approximating the Bayes Optimal Classifier (BOC) using an ensemble of diverse base models with soft voting (Part C).

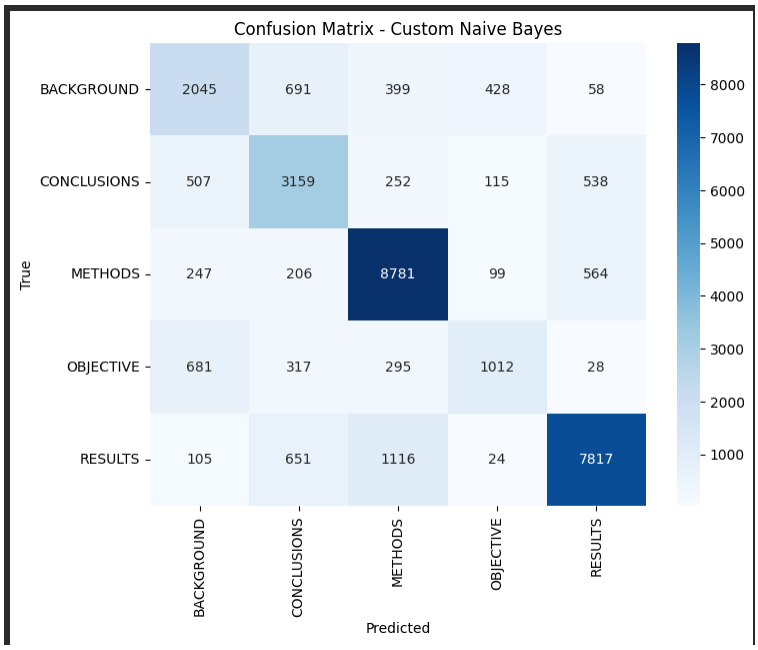
Methodology:

- **MNB Implementation (Part A):**
  Developed a Multinomial Naive Bayes classifier from scratch. Estimated class prior probabilities and class-conditional likelihoods with Laplace smoothing. Utilized a CountVectorizer for token frequency features and predicted labels by aggregating log probabilities for each class.
- **Sklearn Pipeline and GridSearchCV (Part B):**
  Used TfidfVectorizer and MultinomialNB wrapped in a sklearn Pipeline. Performed hyperparameter tuning of n-gram range and smoothing parameter alpha using GridSearchCV with cross-validation on the development dataset to select the best model parameters.
- **Bayes Optimal Classifier (Part C):**
  Approximated BOC by training multiple diverse base classifiers (hypotheses) on bootstrap samples of the training data, computing their posterior weights based on log-likelihoods of data, and combining them in a soft voting ensemble weighted by these posterior probabilities.

Results and Analysis

- **Part A:**

```
=== 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
```

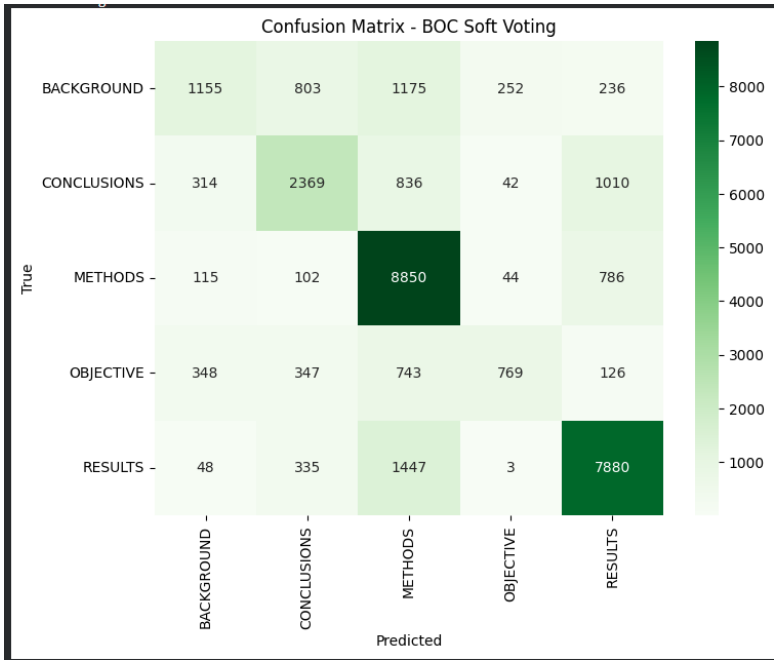


- **Part B:**

```
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...
Grid search complete.
Best parameters: {'nb__alpha': 0.1, 'tfidf__ngram_range': (1, 1)}
Best macro F1 score: 0.5924853482093159
```

- **Part C:**

```
Please enter your full SRN (e.g., PES1UG22CS345): PES2UG23CS338
Using dynamic sample size: 10338
Actual sampled training set size used: 10338

Training all base models...
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247: FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed in 1.7. From then on, it will always use 'multinomial'. Leave
  warnings.warn(
All base models trained.

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

Predicting on test set...

--- Final Evaluation: Bayes Optimal Classifier (Soft Voting) ---
Accuracy: 0.6976
              precision    recall  f1-score   support

  BACKGROUND       0.58      0.32      0.41      3621
 CONCLUSIONS       0.60      0.52      0.56      4571
     METHODS       0.68      0.89      0.77      9897
   OBJECTIVE       0.69      0.33      0.45      2333
     RESULTS       0.79      0.81      0.80      9713

    accuracy                           0.70     30135
   macro avg       0.67      0.57      0.60     30135
weighted avg       0.69      0.70      0.68     30135

Macro-averaged F1 score: 0.5968
```



Discussion:

The custom Count-Based Naive Bayes (Part A) achieved the best overall results with an accuracy of 0.7571 and a macro F1 score of 0.6825. The tuned Sklearn model (Part B), despite hyperparameter tuning, had a lower accuracy (0.6996) and a macro F1 score of 0.5555; its best-tuned F1 on the dev set was 0.5925. Both these models outperformed the Bayes Optimal Classifier (Part C) ensemble, which produced an accuracy of 0.6966 and a macro F1 score of 0.5968.

This suggests that the custom model is well-suited to the dataset, possibly due to effective handling of feature frequencies or regularization. The ensemble BOC approach, while theoretically promising, did not surpass the tailored single-model approach, likely due to limited diversity among base models or insufficient data for robust ensemble weighting. In summary, the scratch Naive Bayes delivered the most balanced and highest performance, followed by the BOC approximation, with the default-tuned Sklearn pipeline trailing behind.

- Custom NB: Highest macro F1 and accuracy—well-optimized for the data.

- Sklearn NB: Moderate improvement with tuning, but less effective than the scratch model.
- BOC (Soft Voting): Did not exceed single model—ensemble averaging was not enough to outperform a strong custom classifier on this task.