

ML Lab

Lab 11 - Submission

Name: Gauthamdev R Holla

SRN: PES2UG23CS197

Branch: CSE

Sem: V

Section: C

Introduction

Focuses on building and evaluating a text classification system using the Multinomial Naive Bayes (MNB) algorithm. The goal is to classify sentences into five categories: BACKGROUND, OBJECTIVE, METHODS, RESULTS, and CONCLUSIONS.

The lab is divided into three parts:

Part A: Implementing MNB from scratch.

Part B: Using Scikit-learn's MNB with TF-IDF features and tuning hyperparameters.

Part C: Approximating the Bayes Optimal Classifier (BOC) using an ensemble of five diverse models.

Methodology

Part A:

- Built a Naive Bayes classifier from scratch using word counts.
- Calculated class probabilities with Laplace smoothing
- Used log probabilities to make predictions.

Part B:

- Used Scikit-learn's *TfidfVectorizer* and *MultinomialNB* in a pipeline.
- Applied *GridSearchCV* to find the best settings for n-grams and smoothing.

Part C:

- Trained 5 different models on a sample of the training data.
- Calculated how well each model performed on a validation set and used those scores to assign weights.

- Combined the models using a soft voting ensemble.

Results and Analysis

Part A:

```

=== Test Set Evaluation (Custom Count-Based Naive Bayes) ===
Accuracy: 0.7571

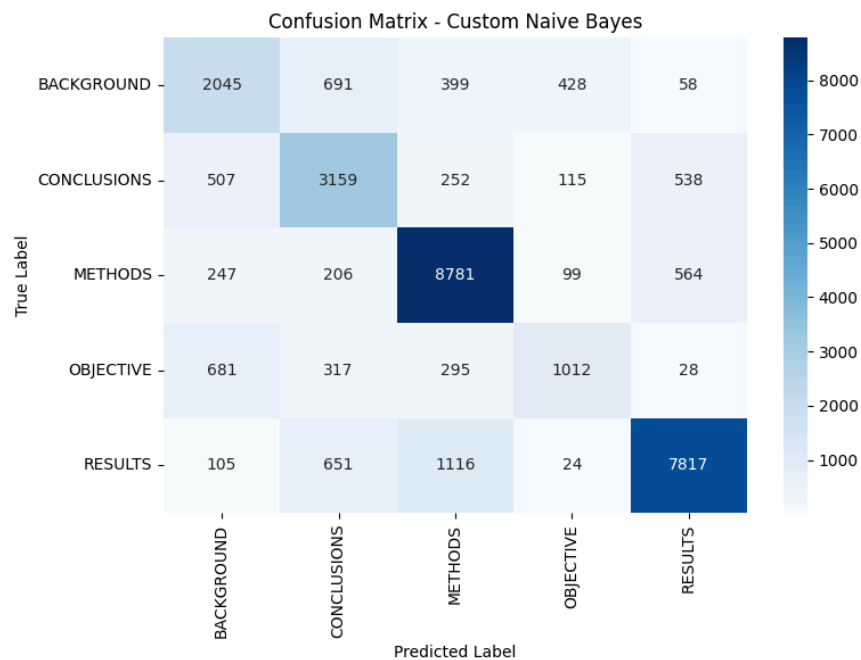
```

	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:

```

=== 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 ===
Best Parameters: {'nb_alpha': 0.1, 'tfidf_ngram_range': (1, 1)}
Best Macro F1 Score: 0.5925

```

Part C:

```
Please enter your full SRN (e.g., PES1UG22CS345): PES2UG23CS197
Using dynamic sample size: 10197
Actual sampled training set size used: 10197

Training all base models...
All base models trained.

Calculating Posterior Weights...
Posterior Weights: [0.22854849 0.2419911 0.21283976 0.17378258 0.14283806]

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

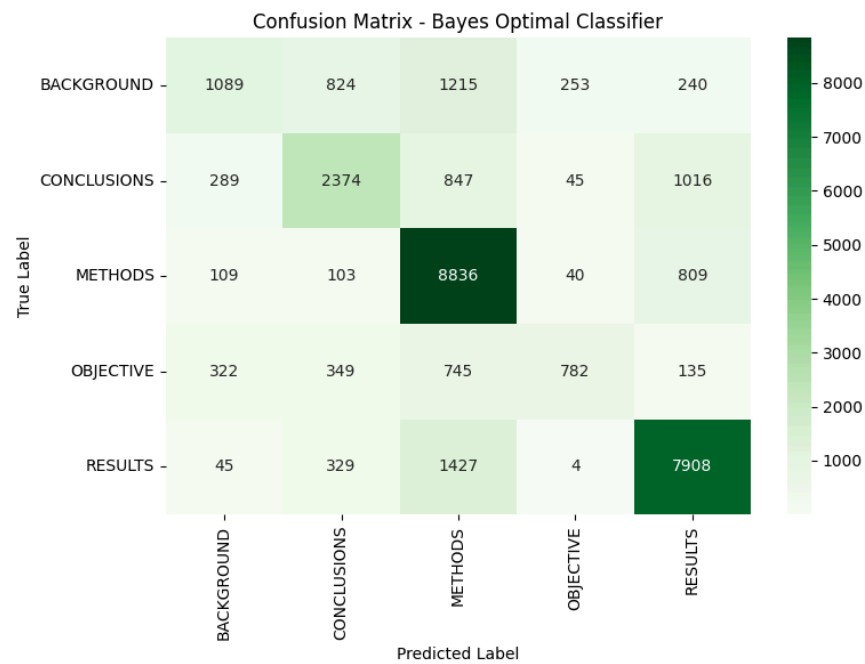
Predicting on test set...
```

```
=== Final Evaluation: Bayes Optimal Classifier (Soft Voting) ===
Accuracy: 0.6965
      precision    recall  f1-score   support

BACKGROUND      0.59      0.30      0.40      3621
CONCLUSIONS   0.60      0.52      0.56      4571
METHODS          0.68      0.89      0.77     9897
OBJECTIVE        0.70      0.34      0.45      2333
RESULTS          0.78      0.81      0.80      9713

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

Macro-averaged F1 score: 0.5946
```



Discussion

Model	Accuracy	Macro F1 Score
Part A: Custom MNB	0.7571	0.6825
Part B: Sklearn MNB	0.6996	0.5555

Part C: BOC Ensemble	0.6965	0.5946
----------------------	--------	--------

- Custom Naive Bayes model (Part A) had the highest scores
 - The Scikit-learn model (Part B) benefited from tuning but didn't generalize as well.
 - The BOC ensemble (Part C) did well in handling diverse sentence types.
-