Custom Naive Bayes Classifier vs. Scikit-learn's Naive Bayes Classifier

Introduction

This report evaluates the performance of a custom Naive Bayes Classifier implemented from scratch compared to the Naive Bayes Classifier provided by the popular machine learning library, Scikit-learn. The evaluation is based on a dataset regarding nursery admissions, where various attributes are considered for the final evaluation of the nursery.

Methodology

1. **Data Preprocessing**: The dataset was loaded from a CSV file and categorical variables were converted into numerical using label encoding.
2. **Model Development**:
    - **Custom Naive Bayes Classifier**: Implemented from scratch, including methods for fitting the model and making predictions.
    - **Scikit-learn's Naive Bayes Classifier**: Utilised GaussianNB from Scikit-learn library as a benchmark.

Results

Classifier	**Accuracy**	**Macro avg F1-score**	**Weighted average F1-score**
Custom Naive Bayes Classifier	66.74%	0.46	0.74
Scikit-learn's Naive Bayes Classifier	75.81%	0.66	0.80

- **Custom Naive Bayes Classifier**:
    - Accuracy: 66.74%
    - Classification Report:

Class	**Precision**	**Recall**	**F1-score**	**Support**
0	0.00	0.00	0.00	0
1	0.89	0.29	0.44	827

2	1.00	0.92	0.96	900
3	0.07	0.21	0.10	58
4	0.88	0.80	0.84	807
Macro avg	0.57	0.44	0.46	2592
Weighted avg	0.91	0.67	0.74	2592

- **Scikit-learn's Naive Bayes Classifier**:
  - Accuracy: 75.81%
  - Classification Report:

Class	**Precision**	**Recall**	**F1-score**	**Support**
1	0.83	0.44	0.58	827
2	1.00	1.00	1.00	900
3	0.10	1.00	0.19	58
4	0.94	0.80	0.86	807
Macro avg	0.72	0.81	0.66	2592
Weighted avg	0.91	0.76	0.80	2592

Conclusion

- **Performance Comparison**: Scikit-learn's Naive Bayes Classifier outperforms the custom implementation in terms of accuracy and overall classification metrics.
- **Recommendation**: Given the significant performance difference, it's advisable to utilise Scikit-learn's implementation for this task.
- **Further Analysis**: Hyperparameter tuning could potentially improve the performance of both models.

Hyperparameter Tuning Results

- *Custom Naive Bayes Classifier*: No hyperparameters were explicitly tuned in the custom implementation.
- *Scikit-learn's Naive Bayes Classifier*: The default hyperparameters provided satisfactory results, achieving an accuracy of 75.81%.