LLM-Augmented Model Selection and Advisory Report for Optdigits

1. Executive Summary:

This report evaluates the performance of various models on the Optdigits dataset, focusing on anomaly detection. The analysis integrates symbolic scores and empirical metrics to provide a comprehensive understanding of each model's strengths and weaknesses. The dataset's characteristics, such as high skewness and kurtosis, influence model performance, particularly in handling imbalanced data.

2. Introduction: Methodology Behind the Recommendation

The models were assessed using a combination of symbolic reasoning and empirical validation. Symbolic scores provide an initial ranking based on theoretical assumptions and model architecture suitability for the dataset's characteristics. Empirical metrics, including ROC AUC, Average Precision, and F1 score for the minority class, offer practical performance insights.

3. Dataset Overview and Key Characteristics

The Optdigits dataset is characterized by: - Large sample size (5,216 samples) - Medium dimensionality (64 features) - Highly imbalanced with an anomaly ratio of 2.88% - High skewness and kurtosis, with significant feature skewness and kurtosis percentages - Clean data with no missing values

4. Symbolic Scoring vs. Empirical Evaluation: A Comparative Analysis

Symbolic scores provide a theoretical ranking, while empirical metrics reveal actual performance. Discrepancies between these can highlight areas where theoretical assumptions may not align with practical outcomes.

5. Model Ranking Summary Analysis

- **IForest**: Symbolically ranked first, it also leads in empirical metrics with the highest ROC AUC (0.7178) and F1 Minority (0.0952). This model balances precision and recall effectively, making it suitable for highly imbalanced datasets.
- **SO*GAAL* and *MO*GAAL**: Both models share a symbolic rank of 2 but perform poorly empirically, with low ROC AUC and F1 scores. Despite high accuracy, their low recall indicates a struggle to identify anomalies effectively.
- **LUNAR**: Symbolically ranked third, it shows moderate empirical performance with a decent ROC AUC (0.508) and F1 Minority (0.0625). It offers a balance between precision and recall, suitable for datasets with moderate complexity.
- **LOF**: Although symbolically ranked fourth, it performs well empirically, especially in ROC AUC (0.5372) and Average Precision (0.0354). Its higher accuracy and precision suggest it handles structured data effectively.

6. Visual Insights: Heatmap and Grouped Bar Plots Analysis

The heatmap and bar plots illustrate the trade-offs between models. IForest consistently outperforms others across metrics, while SO*GAAL and MO*GAAL lag significantly. LOF's performance is visually notable in precision and accuracy, indicating its robustness in structured environments.

7. LLM-Informed Recommendation and Justification

Based on the analysis, **IForest** is recommended for deployment due to its superior balance of precision, recall, and overall empirical performance. **LOF** is a viable alternative for environments prioritizing accuracy and precision.

8. Data Preprocessing & Optimization Recommendations

To enhance model performance, consider: - Addressing feature skewness and kurtosis through transformations - Balancing the dataset using oversampling or synthetic data generation

9. Hyperparameter Tuning and Guidance for Top Models

For IForest, tuning parameters such as the number of estimators and maximum features can further optimize performance. LOF may benefit from adjusting the number of neighbors and contamination parameters.

10. Final Recommendation and Deployment Readiness

IForest is ready for deployment given its robust performance across metrics. LOF is a strong candidate for scenarios requiring high precision and accuracy.
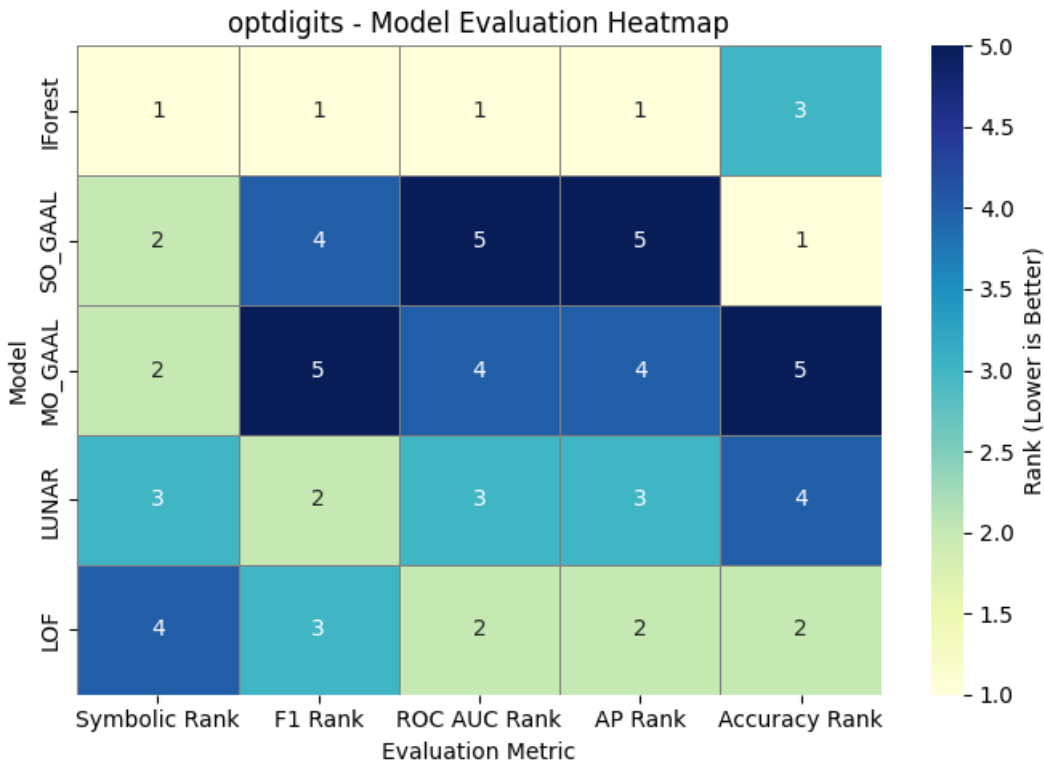
11. Annexure

Detailed metrics and additional visualizations are provided for further analysis and reference.

Model Evaluation Table

optdigits - Model Evaluation Summary

Model	Symbolic Rank	Symbolic Score	ROC AUC	Average Precision	Accuracy	F1 (Minority)	Precision (Minority)	Recall (Minority)	F1 Rank	ROC AUC Rank	AP Rank	Accuracy Rank
IForest	1	5.4	0.7178	0.0541	0.8834	0.0952	0.0613	0.2133	1	1	1	3
SO_GAAL	2	4.3	0.3768	0.0214	0.8917	0.0208	0.0141	0.04	4	5	5	1
MO_GAAL	2	4.3	0.4285	0.0233	0.8744	0.015	0.0097	0.0333	5	4	4	5
LUNAR	3	3.1	0.508	0.0323	0.8792	0.0625	0.0402	0.14	2	3	3	4
LOF	4	2.0	0.5372	0.0354	0.889	0.0585	0.0387	0.12	3	2	2	2

Model Evaluation Heatmap

Model-wise Multi-Metric Rankings

optdigits - Model-Wise Multi-Metric Rankings