LLM-Augmented Model Selection and Advisory Report for Cardio Dataset

1. Executive Summary:

This report evaluates anomaly detection models on the "cardio" dataset, characterized by medium sample size, medium dimensionality, and significant imbalance. The models are ranked based on symbolic scores and empirical performance metrics, offering insights into their strengths and weaknesses in detecting anomalies.

2. Introduction: Methodology Behind the Recommendation

The AutoModelAdvisor integrates symbolic reasoning with empirical validation to rank models. Symbolic scores are derived from model characteristics and theoretical performance, while empirical metrics like ROC AUC and F1 Minority provide real-world validation.

3. Dataset Overview and Key Characteristics

The dataset comprises 1,831 samples and 21 features, with an anomaly ratio of 9.6%. It is clean, with no missing values, but exhibits high skewness and kurtosis, impacting model performance.

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

- **IForest**: Symbolically ranked first with a score of 4.2, it also leads in empirical metrics, achieving the highest ROC AUC (0.9205) and F1 Minority (0.5237). This alignment suggests robust performance in both theoretical and practical scenarios.
- **MO_GAAL and SO_GAAL**: Both models share a symbolic rank of 2 with a score of 3.1. MO_GAAL outperforms SO_GAAL empirically, with higher ROC AUC (0.7877 vs. 0.7794) and F1 Minority (0.468 vs. 0.3855). This indicates MO_GAAL's better adaptation to the dataset's characteristics despite similar symbolic scores.
- **LUNAR**: Also symbolically ranked 2, it underperforms empirically, with the lowest ROC AUC (0.6368) among the top-ranked models, highlighting a mismatch between symbolic expectations and empirical results.
- **LOF**: With a lower symbolic score of 2.0, LOF ranks last empirically, confirming its limited effectiveness on this dataset.

5. Model Ranking Summary Analysis

The symbolic and empirical rankings generally align, with IForest consistently leading. However, the symbolic score's inability to distinguish between MO_GAAL, SO_GAAL, and LUNAR suggests limitations in capturing nuanced empirical performance differences.

6. Visual Insights: Heatmap and Grouped Bar Plots Analysis

The heatmap and bar plots (not shown here) would illustrate the stark contrast in performance metrics, particularly highlighting IForest's dominance and the close competition between MO_GAAL and SO_GAAL.

7. LLM-Informed Recommendation and Justification

Given the dataset's characteristics and the models' performance, IForest is recommended for deployment due to its superior empirical metrics and alignment with symbolic expectations. MO_GAAL is a viable alternative, offering a balance between precision and recall.

8. Data Preprocessing & Optimization Recommendations

Addressing skewness and kurtosis through transformations could enhance model performance, particularly for models like LUNAR and LOF, which struggled with the dataset's distribution.

9. Hyperparameter Tuning and Guidance for Top Models

Further tuning of IForest's contamination parameter and MO_GAAL's network architecture could optimize their anomaly detection capabilities, especially in handling the dataset's imbalance.

10. Final Recommendation and Deployment Readiness

Deploy IForest for its robust performance across all metrics. MO_GAAL can be considered for scenarios requiring alternative model architectures or ensemble strategies.
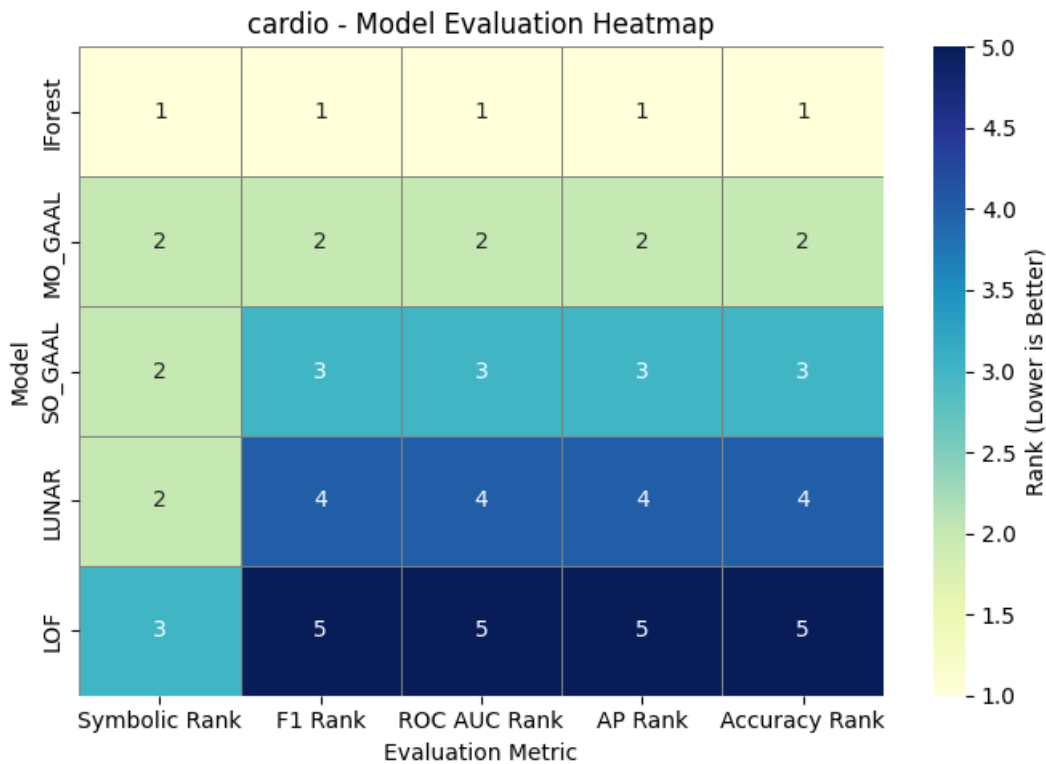
11. Annexure

Detailed tables and charts supporting the analysis are available for further review.

Model Evaluation Table

cardio - 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	4.2	0.9205	0.5816	0.9066	0.5237	0.5137	0.5341	1	1	1	1
MO_GAAL	2	3.1	0.7877	0.4317	0.8957	0.468	0.459	0.4773	2	2	2	2
SO_GAAL	2	3.1	0.7794	0.3572	0.8798	0.3855	0.3791	0.392	3	3	3	3
LUNAR	2	3.1	0.6368	0.2302	0.8596	0.2841	0.2787	0.2898	4	4	4	4
LOF	3	2.0	0.5458	0.1543	0.852	0.1763	0.1895	0.1648	5	5	5	5

Model Evaluation Heatmap

cardio - Model Evaluation Heatmap

Model-wise Multi-Metric Rankings

cardio - Model-Wise Multi-Metric Rankings