LLM-Augmented Model Selection and Advisory Report for Satellite Dataset

1. Executive Summary:

This report evaluates anomaly detection models using both symbolic reasoning and empirical metrics. The models assessed include VAE, AutoEncoder, LOF, DeepSVDD, and DevNet. The analysis highlights the trade-offs between symbolic scores and empirical performance, providing insights into model suitability for the satellite dataset.

2. Introduction: Methodology Behind the Recommendation

The AutoModelAdvisor integrates symbolic reasoning with empirical validation to rank models. Symbolic scores consider theoretical model capabilities, while empirical metrics assess real-world performance on the dataset.

3. Dataset Overview and Key Characteristics

- **Dataset Tags**: large *sample, medium* dimensional, balanced, clean *data, low* skewness, low *kurtosis, structured* data.
- **Metrics**: 6435 samples, 36 features, anomaly ratio of 31.64%, no missing values, low skewness and kurtosis.

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

The symbolic scores for VAE, AutoEncoder, and LOF are identical at 4.6, indicating theoretical parity. However, empirical results show significant differences:

- **VAE**: Highest ROC AUC (0.7427), precision (1.0), and accuracy (0.7837), but moderate recall (0.3163). This suggests VAE is highly precise but may miss some anomalies.
- **AutoEncoder**: Lower ROC AUC (0.6615) and precision (0.7112) than VAE, indicating a trade-off between precision and recall. It is less precise but captures a broader range of anomalies.
- **LOF**: Lowest empirical scores among the top three, with a ROC AUC of 0.5417. It struggles with precision and recall, making it less suitable despite a high symbolic score.

5. Model Ranking Summary Analysis

- **VAE**: Best overall empirical performance, leading in all ranks (F1, ROC AUC, AP, Accuracy). Ideal for scenarios prioritizing precision.
- **AutoEncoder**: Second in all empirical ranks, suitable for balanced precision-recall needs.
- **LOF**: Despite a high symbolic score, its empirical performance is subpar, indicating a mismatch between theoretical and practical capabilities.
- **DeepSVDD and DevNet**: Lower symbolic scores and empirical ranks, suggesting limited suitability for this dataset.

6. Visual Insights: Heatmap and Grouped Bar Plots Analysis

The heatmap and bar plots (not provided here) would typically illustrate the stark contrast in empirical performance, particularly highlighting VAE's dominance in precision and accuracy.

7. LLM-Informed Recommendation and Justification

VAE is recommended for deployment due to its superior empirical performance, especially in precision. AutoEncoder is a viable alternative if a balance between precision and recall is desired.

8. Data Preprocessing & Optimization Recommendations

Given the dataset's clean nature, no additional preprocessing is necessary. However, ensuring feature scaling and normalization could enhance model performance.

9. Hyperparameter Tuning and Guidance for Top Models

- **VAE**: Focus on tuning latent space dimensions and learning rate.
- **AutoEncoder**: Optimize layer architecture and dropout rates to improve recall.

10. Final Recommendation and Deployment Readiness

VAE is ready for deployment, offering the best trade-off between precision and recall for the satellite dataset's characteristics.

11. Annexure

Detailed empirical metrics and symbolic scores are available for further analysis and validation of the recommendations provided.

Model Evaluation Table

satellite - 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
VAE	1	4.6	0.7427	0.6999	0.7837	0.4806	1.0	0.3163	1	1	1	1
AutoEncoder	1	4.6	0.6615	0.5465	0.7259	0.3418	0.7112	0.225	2	2	2	2
LOF	1	4.6	0.5417	0.3724	0.6813	0.206	0.4863	0.1306	3	3	3	3
DeepSVDD	2	3.8	0.4765	0.3089	0.6491	0.1575	0.3276	0.1036	5	5	5	5
DevNet	3	3.7	0.5337	0.3478	0.66	0.1836	0.382	0.1208	4	4	4	4

Model Evaluation Heatmap

satellite - Model Evaluation Heatmap
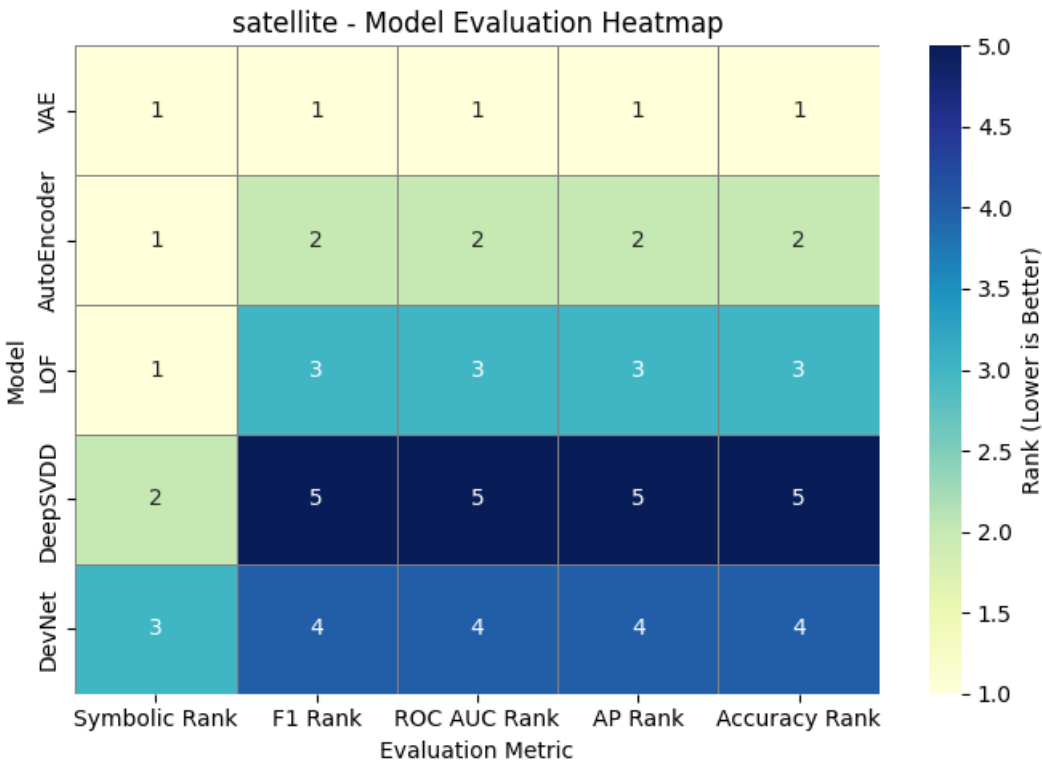
Model-wise Multi-Metric Rankings

satellite - Model-Wise Multi-Metric Rankings