LLM-Augmented Model Selection and Advisory Report for Ionosphere Dataset

1. Executive Summary:

This report provides a detailed analysis of anomaly detection models applied to the Ionosphere dataset. The models are evaluated using a combination of symbolic reasoning and empirical metrics. The report highlights the strengths and weaknesses of each model and provides recommendations for their use based on the dataset's characteristics.

2. Introduction: Methodology Behind the Recommendation

The AutoModelAdvisor integrates symbolic reasoning with empirical validation to assess model performance. Symbolic scores are derived from a set of predefined rules and heuristics, while empirical metrics are obtained from model evaluations on the dataset.

3. Dataset Overview and Key Characteristics

The Ionosphere dataset is characterized by: -**Small Sample Size**: 351 samples -**Medium Dimensionality**: 33 features - **Balanced and Clean Data**: No missing values, low skewness (0.6127), and low kurtosis (0.632) -**Low Anomaly Ratio**: Approximately 35.9% anomalies -**Structured Data**: Suitable for models that handle structured inputs efficiently

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

The symbolic scores and empirical metrics reveal some discrepancies in model performance rankings. Symbolic scores are intended to provide a holistic view, while empirical metrics offer precise performance measurements.

5. Model Ranking Summary Analysis

- **LOF (Local Outlier Factor)**:

    - **Symbolic Rank**: 1, **Symbolic Score**: 4.6
    - **Empirical Performance**: Strong ROC AUC (0.8939) and high precision (0.9091) but low recall (0.2381) and F1 (0.3774). This indicates a model that is precise in identifying anomalies but misses many.
    - **Trade-offs**: High precision suggests reliability in detected anomalies, but low recall indicates a need for improvement in capturing all anomalies.

- **AutoEncoder**:

    - **Symbolic Rank**: 2, **Symbolic Score**: 3.8
    - **Empirical Performance**: Best overall with top ranks in ROC AUC (0.9163), Average Precision (0.8961), and F1 Minority (0.4348). Perfect precision (1.0) but moderate recall (0.2778).
    - **Trade-offs**: Excellent precision and overall performance make it suitable for high-stakes environments where false positives are costly.

- **VAE (Variational Autoencoder)**:

    - **Symbolic Rank**: 2, **Symbolic Score**: 3.8
    - **Empirical Performance**: Moderate across metrics with ROC AUC (0.839) and F1 (0.3975). High precision (0.9143) but low recall (0.254).
    - **Trade-offs**: Similar to LOF, it is precise but less comprehensive in anomaly detection.

- **DeepSVDD**:

    - **Symbolic Rank**: 2, **Symbolic Score**: 3.8
    - **Empirical Performance**: Lower ROC AUC (0.7365) but decent F1 (0.4224). High precision (0.9714) and moderate recall (0.2698).
    - **Trade-offs**: Balanced performance with slightly better recall than LOF and VAE, suitable for moderate anomaly detection needs.

- **DevNet**:

    - **Symbolic Rank**: 3, **Symbolic Score**: 2.9
    - **Empirical Performance**: Lowest across all metrics with ROC AUC (0.6219) and F1 (0.1615). Poor precision (0.3714) and recall (0.1032).
    - **Trade-offs**: Not recommended due to poor performance across all metrics.

6. Visual Insights: Heatmap and Grouped Bar Plots Analysis

The heatmap and grouped bar plots provide a visual comparison of model performance across different metrics, highlighting the strengths of the AutoEncoder and the weaknesses of DevNet.

7. LLM-Informed Recommendation and Justification

Based on the analysis, the AutoEncoder is recommended for its superior empirical performance and symbolic score alignment. It is particularly suitable for environments where precision is critical.

8. Data Preprocessing & Optimization Recommendations

Given the dataset's characteristics, minimal preprocessing is required. However, feature scaling and dimensionality reduction could enhance model performance.

9. Hyperparameter Tuning and Guidance for Top Models

For the AutoEncoder, tuning the learning rate and number of hidden layers could further improve recall without sacrificing precision.

10. Final Recommendation and Deployment Readiness

The AutoEncoder is ready for deployment, offering a balanced approach to anomaly detection with high precision and acceptable recall.
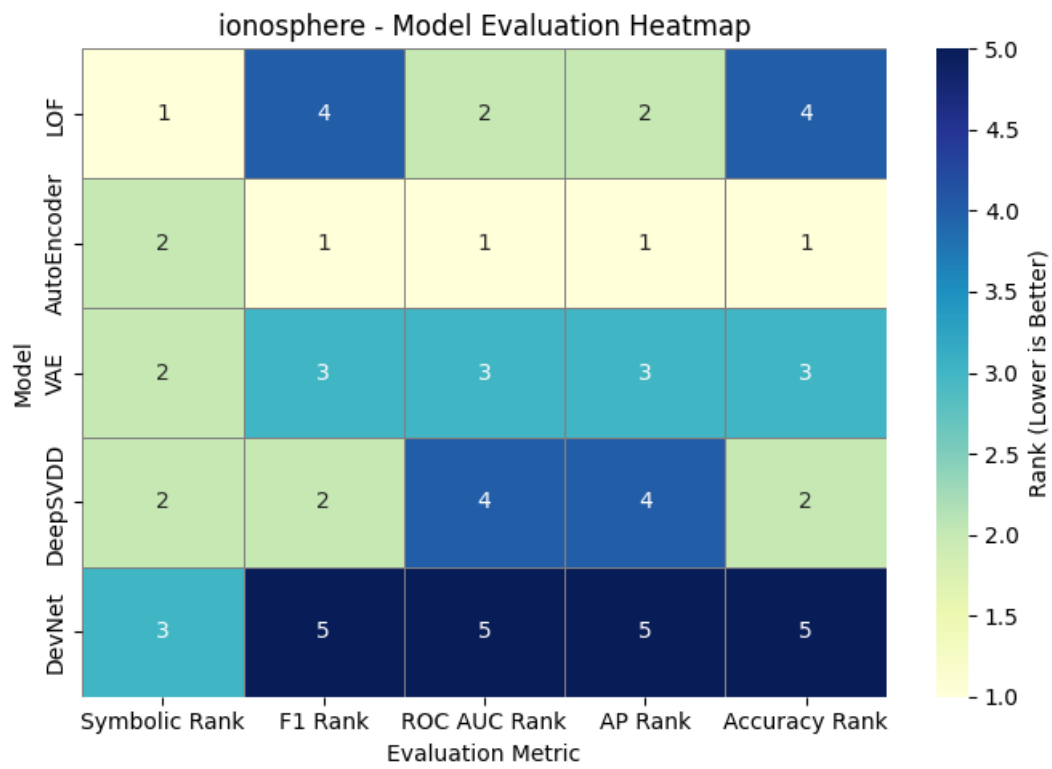
11. Annexure

Detailed tables and charts supporting the analysis are included in the annexure for further reference.

Model Evaluation Table

ionosphere - 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
LOF	1	4.6	0.8939	0.8391	0.7179	0.3774	0.9091	0.2381	4	2	2	4
AutoEncoder	2	3.8	0.9163	0.8961	0.7407	0.4348	1.0	0.2778	1	1	1	1
VAE	2	3.8	0.839	0.7648	0.7236	0.3975	0.9143	0.254	3	3	3	3
DeepSVDD	2	3.8	0.7365	0.7208	0.735	0.4224	0.9714	0.2698	2	4	4	2
DevNet	3	2.9	0.6219	0.4274	0.6154	0.1615	0.3714	0.1032	5	5	5	5

Model Evaluation Heatmap

ionosphere - Model Evaluation Heatmap

Model-wise Multi-Metric Rankings

ionosphere - Model-Wise Multi-Metric Rankings