LLM-Augmented Model Selection and Advisory Report for Pendigits Dataset

1. Executive Summary:

The anomaly detection task on the Pendigits dataset, characterized by a large sample size, medium dimensionality, and high imbalance, was evaluated using several models. The models were ranked based on symbolic scores and empirical metrics, revealing both alignments and discrepancies. The Variational Autoencoder (VAE) and Isolation Forest (IForest) emerged as top performers under different conditions.

2. Introduction: Methodology Behind the Recommendation

The AutoModelAdvisor integrates symbolic reasoning, empirical validation, and LLM guidance to evaluate models. Symbolic scores reflect theoretical suitability, while empirical metrics provide performance validation.

3. Dataset Overview and Key Characteristics

- **Sample Size**: 6870
- **Features**: 16
- **Anomaly Ratio**: 2.27%
- **Data Quality**: Clean, with no missing values
- **Distribution**: Low skewness and kurtosis

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

The symbolic score, a theoretical measure of model suitability, ranked VAE, AutoEncoder, LOF, and DeepSVDD equally at 3.8, while IForest scored 3.2. However, empirical metrics reveal significant performance differences:

- **VAE**: Achieved the highest ROC AUC (0.9469) and F1 Minority (0.2732), indicating strong detection capability and balance between precision and recall.
- **IForest**: Close to VAE in ROC AUC (0.9464) but superior in Average Precision (0.2944), suggesting better precision in identifying anomalies.
- **AutoEncoder**: Lower ROC AUC (0.8416) and F1 Minority (0.1139), indicating weaker performance in detecting anomalies.
- **LOF and DeepSVDD**: Both models underperformed with low ROC AUC and F1 Minority scores, reflecting limited anomaly detection capability.

5. Model Ranking Summary Analysis

- **VAE**: Best overall empirical performance, excelling in ROC AUC and F1 Minority, making it suitable for scenarios prioritizing balanced detection.
- **IForest**: Best Average Precision, ideal for applications where minimizing false positives is critical.
- **AutoEncoder, LOF, DeepSVDD**: Lower empirical performance, suggesting limited applicability for this dataset.

6. Visual Insights: Heatmap and Grouped Bar Plots Analysis

The heatmap and bar plots illustrate the stark contrast between symbolic scores and empirical metrics, highlighting the importance of empirical validation in model selection.

7. LLM-Informed Recommendation and Justification

For the Pendigits dataset, VAE is recommended for its balanced performance across metrics, while IForest is suggested for applications requiring high precision. The symbolic score alignment with empirical results for VAE and IForest supports their selection.

8. Data Preprocessing & Optimization Recommendations

Given the clean nature of the dataset, focus on feature scaling and dimensionality reduction to enhance model performance.

9. Hyperparameter Tuning and Guidance for Top Models

- **VAE**: Experiment with latent space dimensionality and learning rate adjustments.
- **IForest**: Optimize the number of trees and sample size for improved precision.

10. Final Recommendation and Deployment Readiness

Deploy VAE for balanced anomaly detection and IForest for precision-critical tasks. Ensure continuous monitoring and retraining to adapt to potential data shifts.

11. Annexure

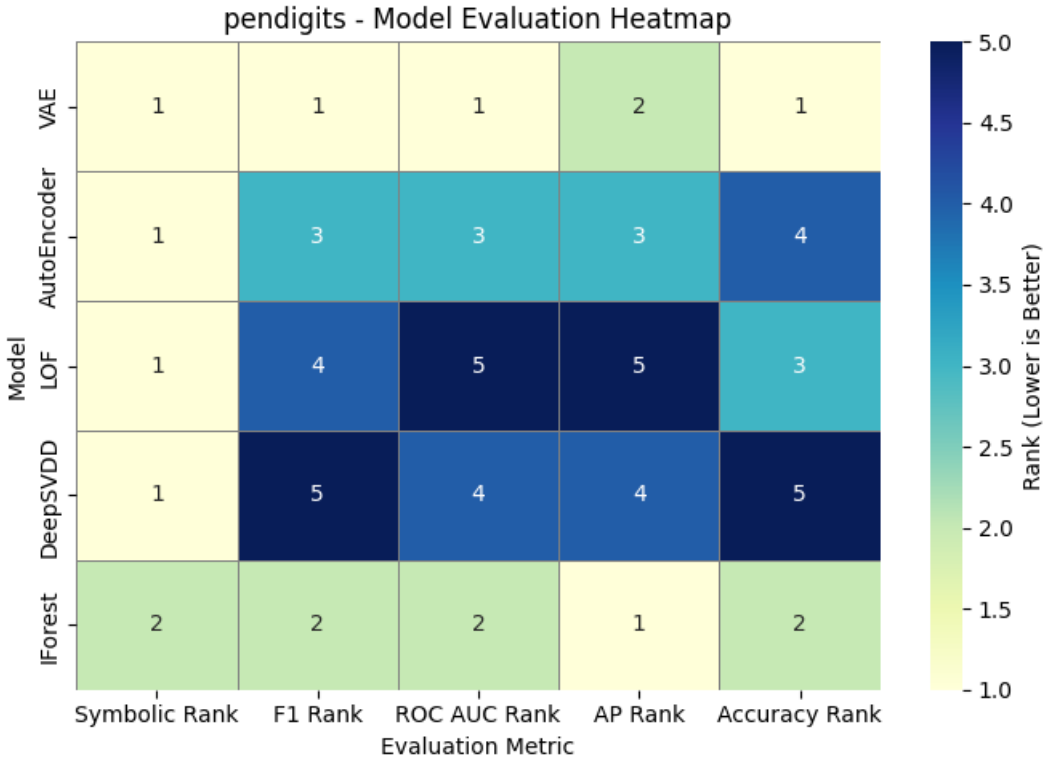
Detailed metric tables, heatmaps, and bar plots are provided for further analysis and validation.

Model Evaluation Table

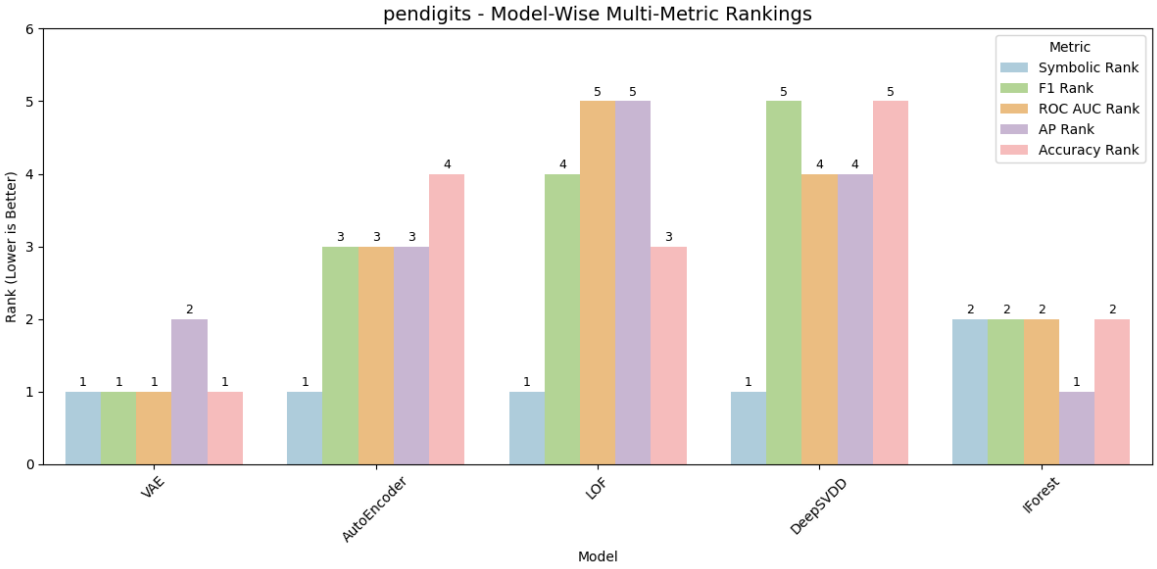
pendigits - Model Evaluation Summary

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	3.8	0.9469	0.2578	0.9109	0.2732	0.1676	0.7372	1	1	2	1
AutoEncoder	1	3.8	0.8416	0.0797	0.8913	0.1139	0.0699	0.3077	3	3	3	4
LOF	1	3.8	0.4991	0.0431	0.8968	0.0732	0.046	0.1795	4	5	5	3
DeepSVDD	1	3.8	0.7532	0.0457	0.8834	0.0498	0.0306	0.1346	5	4	4	5
IForest	2	3.2	0.9464	0.2944	0.9102	0.2681	0.1645	0.7244	2	2	1	2

Model Evaluation Heatmap



Model-wise Multi-Metric Rankings


pendigits - Model-Wise Multi-Metric Rankings
Metric
Symbolic Rank
F1 Rank
ROC AUC Rank
AP Rank
Accuracy Rank
Rank (Lower is Better)
Model
VAE
AutoEncoder
LOF
DeepSVDD
iForest