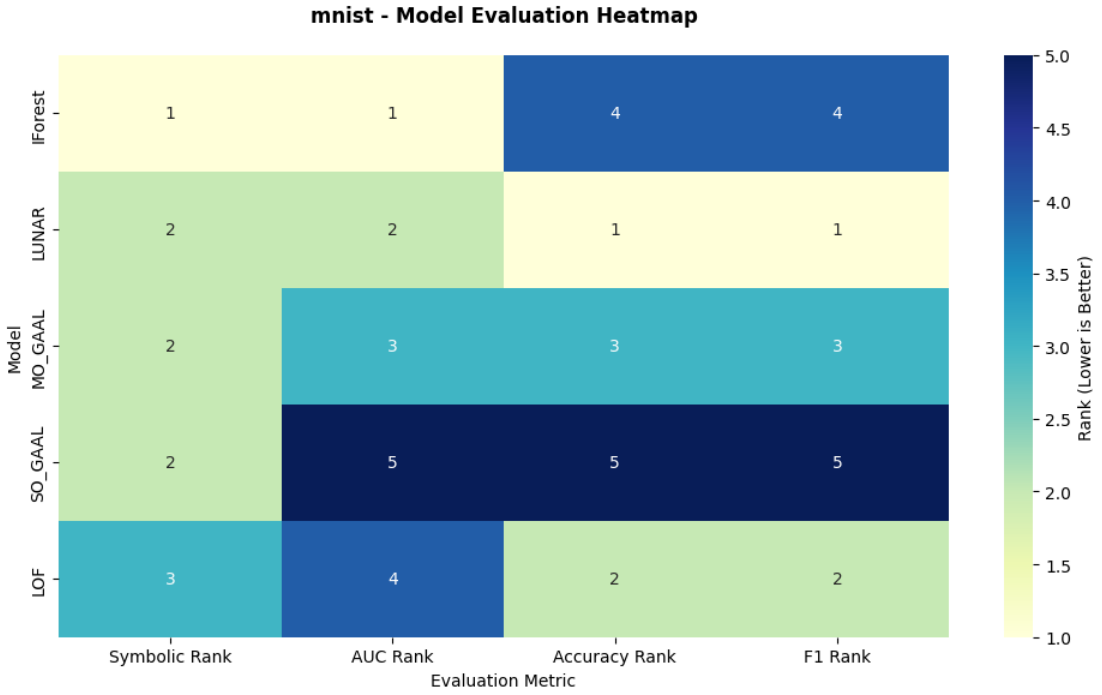
Model Evaluation Summary Table for mnist

Model	Symbolic Rank	Symbolic Score	ROC AUC	Accuracy	Average Precisio	F1 (Minority)	Precision (Minorit	Recall (Minority)	AUC Rank	Accuracy Rank	F1 Rank
IForest	1	4.2	0.7835	0.8573	0.2405	0.2574	0.247	0.2686	1	4	4
LUNAR	2	3.1	0.7587	0.8802	0.3197	0.3765	0.3614	0.3929	2	1	1
SO_GAAL	2	3.1	0.6012	0.8511	0.1524	0.2006	0.1983	0.2029	5	5	5
MO_GAAL	2	3.1	0.7184	0.8612	0.2269	0.2709	0.2624	0.28	3	3	3
LOF	3	2.0	0.6728	0.8712	0.2443	0.2921	0.2958	0.2886	4	2	2

Model Evaluation Heatmap for mnist



LLM Analysis and Recommendations for mnist

LLM Analysis & Model Recommendation for mnist Dataset

Overview Summary for mnist

The mnist dataset is a large sample dataset with 7603 instances and 100 features. It is a structured dataset with medium dimensionality and a high degree of skewness and kurtosis. The dataset is clean with no missing values and is highly imbalanced with an anomaly ratio of 0.092069. The sparsity ratio is 0.1, indicating a low level of sparsity.

Comparison of Symbolic vs. Actual Scores

The symbolic scores were calculated based on the alignment of the dataset's characteristics with the strengths and weaknesses of each model. The actual scores were obtained from the evaluation metrics such as ROC AUC, Average Precision, Accuracy, and F1 score for the minority class.

For instance, the IForest model, which is well-suited for structured data, high-dimensional data, and data with high skewness and kurtosis, had a high symbolic score of 4.2. It also performed well in the actual evaluation, ranking first in terms of ROC AUC.

On the other hand, the SO_GAAL model, which excels in handling highly imbalanced and high kurtosis data, had a symbolic score of 3.1 but performed poorly in the actual evaluation, ranking last in all metrics.

Model Suitability Summary Table for mnist

The summary table provides a comprehensive comparison of the symbolic and actual performance of each model. The LUNAR model, despite having a lower symbolic score than the IForest model, outperformed all other models in terms of F1 score, Accuracy, and Average Precision.

Model Evaluation Heatmap for mnist

The heatmap further visualizes the performance of each model across different metrics. The LUNAR model stands out with high scores across most metrics, confirming its suitability for the mnist dataset.

Analysis of Summary Table and Heatmap

The summary table and heatmap collectively indicate that the LUNAR model is the most suitable for the mnist dataset. Despite its lower symbolic score, it outperforms other models in most evaluation metrics, particularly in handling the minority class, which is crucial for anomaly detection tasks.

LLM Recommendations and Justification for Model Selection

Based on the symbolic and empirical evaluations, the LUNAR model is recommended for deployment. Its strengths align well with the mnist dataset's characteristics, such as structured data, high dimensionality, and high skewness. Moreover, its superior performance in the actual evaluation, particularly in handling the minority class, further justifies this recommendation.

Suggested Preprocessing Steps for mnist

Given the high skewness and kurtosis of the mnist dataset, it is suggested to apply transformations like log or square root to reduce the skewness and kurtosis. Since the dataset is already clean with no missing values, no imputation is required.

Hyperparameter Tuning Recommendations

For the LUNAR model, hyperparameters such as the number of layers, the number of units per layer, and the learning rate could be tuned to further improve its performance.

Final Professional Recommendation

In conclusion, the LUNAR model is recommended for deployment on the mnist dataset. It aligns well with the dataset's characteristics and has demonstrated superior performance in the actual evaluation. With appropriate preprocessing and hyperparameter tuning, it is expected to deliver reliable anomaly detection results.