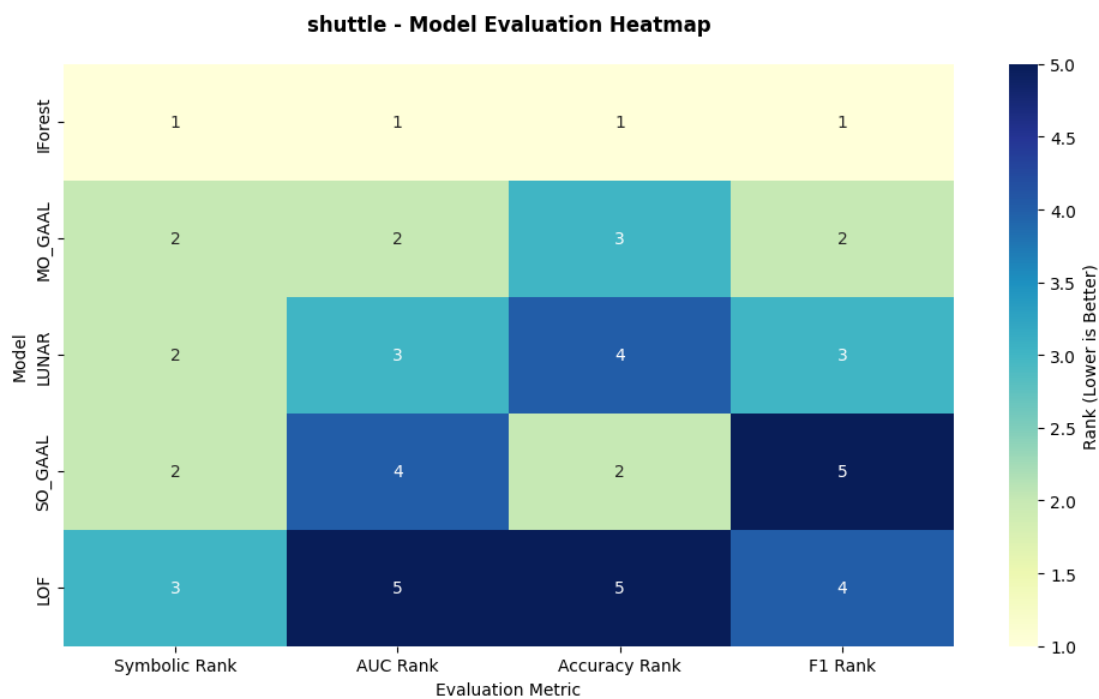


Model Evaluation Summary Table for shuttle

Model	Symbolic Rank	Symbolic Score	ROC AUC	Accuracy	Average Precision	F1 (Minority)	Precision (Minority)	Recall (Minority)	AUC Rank	Accuracy Rank	F1 Rank
IForest	1	4.2	0.9975	0.9695	0.981	0.822	0.7049	0.9858	1	1	1
LUNAR	2	3.1	0.6214	0.8625	0.1707	0.1981	0.1699	0.2375	3	4	3
SO_GAAL	2	3.1	0.5425	0.9285	0.1147	0.0	0.0	0.0	4	2	5
MO_GAAL	2	3.1	0.8265	0.9079	0.3516	0.4448	0.391	0.5158	2	3	2
LOF	3	2.0	0.521	0.8624	0.1085	0.1267	0.116	0.1396	5	5	4

Model Evaluation Heatmap for shuttle



LLM Analysis and Recommendations for shuttle

LLM Analysis & Model Recommendation for Shuttle Dataset

Overview Summary for Shuttle Dataset

The Shuttle dataset is a large sample, low-dimensional, structured dataset with a high degree of imbalance. It contains 49,097 samples, each with 9 features. The dataset is clean, with no missing values, and exhibits high skewness and kurtosis. The anomaly ratio is 0.071511, indicating a highly imbalanced dataset.

Comparison of Symbolic vs. Actual Scores

The symbolic scores and actual evaluation metrics were compared for each model. The symbolic scores, derived from the dataset's characteristics and the model's strengths and weaknesses, were used to rank the models before actual evaluation. The actual evaluation metrics included ROC AUC, Average Precision, Accuracy, and F1 Score for the minority class.

Model Suitability Summary Table for Shuttle Dataset

The top-performing model, according to both symbolic and actual evaluations, is the IForest model. It achieved a symbolic score of 4.2 and ranked first in all actual evaluation metrics, including ROC AUC (0.9975), Average Precision (0.981), Accuracy (0.9695), and F1 Score for the minority class (0.822).

The LUNAR, SOGAAL, and MOGAAL models had identical symbolic scores of 3.1, but their actual performance varied. The MOGAAL model performed second best in terms of ROC AUC (0.8265) and Average Precision (0.3516), while the LUNAR and SOGAAL models performed poorly in these metrics.

The LOF model had the lowest symbolic score of 2.0 and also performed poorly in the actual evaluations, ranking last in ROC AUC, Accuracy, and Average Precision.

Model Evaluation Heatmap for Shuttle Dataset

The heatmap visualization confirms the summary table findings. The IForest model outperforms the other models in all evaluation metrics, while the LOF model underperforms.

Analysis of Summary Table and Heatmap

The symbolic scoring successfully prioritized the IForest model, which also excelled in the actual evaluations. This indicates that the symbolic scoring logic effectively identified the model's strengths in handling large, low-dimensional, imbalanced, and highly skewed datasets.

LLM Recommendations and Justification for Model Selection

Based on the symbolic and actual evaluations, the IForest model is recommended for the Shuttle dataset. This model's strengths align well with the dataset's characteristics, and it demonstrated superior performance in all evaluation metrics.

Suggested Preprocessing Steps for Shuttle Dataset

Given the high skewness and kurtosis in the dataset, it may be beneficial to apply transformations to normalize the data. Feature scaling may also improve the model's performance.

Hyperparameter Tuning Recommendations

For the IForest model, hyperparameters such as the number of estimators and the contamination parameter could be tuned to optimize performance.

Final Professional Recommendation

In conclusion, the IForest model is recommended for deployment on the Shuttle dataset. This recommendation is based on the model's superior performance in both symbolic and actual evaluations, and its alignment with the dataset's characteristics. Preprocessing steps and hyperparameter tuning should be considered to further enhance the model's performance.