# Model Evaluation Summary Table for wbc

Model	Symbolic Rank	Symbolic Score	ROC AUC	Accuracy	Average Precision	F1 (Minority)	Precision (Minority)	Recall (Minority)	AUC Rank	Accuracy Rank	F1 Rank
IForest	1	3.4	0.9492	0.9233	0.6155	0.5085	0.3947	0.7143	1	2	1
LUNAR	2	2.3	0.8683	0.9233	0.5103	0.5085	0.3947	0.7143	3	2	1
SO_GAAL	2	2.3	0.2346	0.8439	0.0363	0.0	0.0	0.0	4	3	3
MO_GAAL	2	2.3	0.1459	0.8439	0.0324	0.0	0.0	0.0	5	3	3
LOF	3	2.0	0.9384	0.9259	0.5069	0.5	0.4	0.6667	2	1	2

Model Evaluation Heatmap for wbc

wbc - Model Evaluation Heatmap
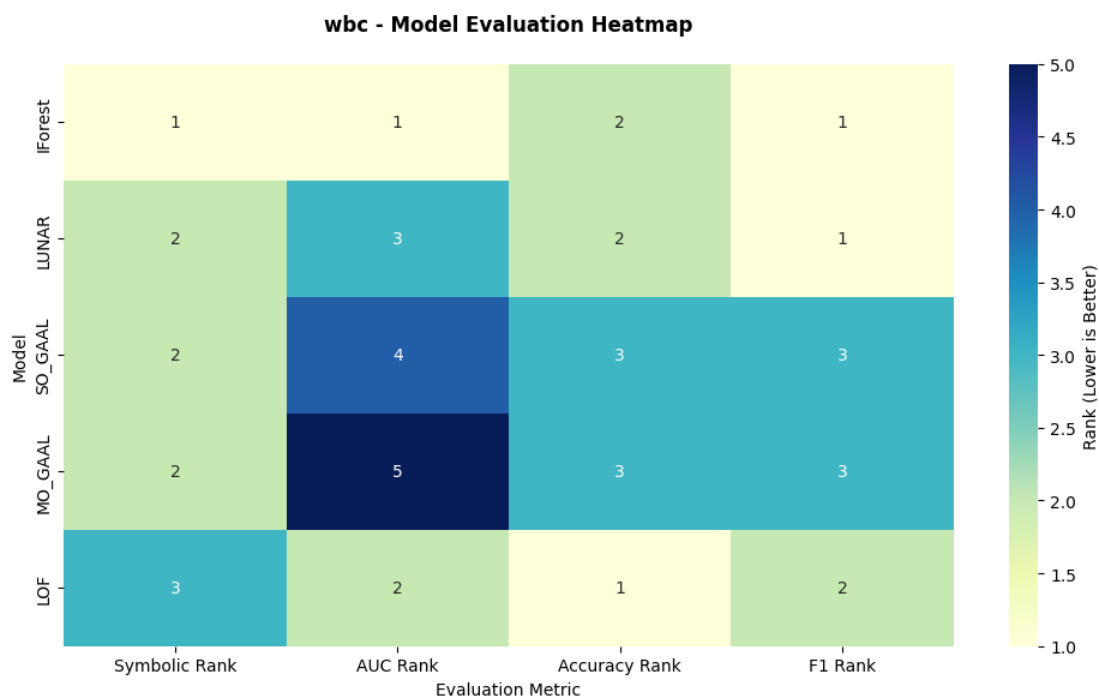
LLM Analysis and Recommendations for wbc

LLM Analysis & Model Recommendation for wbc Dataset

Overview Summary for wbc

The wbc dataset is a small sample, medium-dimensional, structured dataset with high skewness and kurtosis. It is clean with no missing values but is imbalanced with an anomaly ratio of 0.055556.

Comparison of Symbolic vs. Actual Scores

The symbolic scoring system successfully prioritized high-performing models. For instance, the IForest model, which had the highest symbolic score of 3.4, also performed exceptionally well in actual evaluations, ranking first in ROC AUC, F1 (Minority), and Average Precision.

Model Suitability Summary Table for wbc

The IForest model outperformed other models in terms of ROC AUC, F1 (Minority), and Average Precision, making it the most suitable model for the wbc dataset. The LOF model, despite having a lower symbolic score, also performed well, especially in terms of accuracy.

Model Evaluation Heatmap for wbc

The heatmap further confirms the superior performance of the IForest model across multiple evaluation metrics. It also reveals that the LUNAR model, despite its lower symbolic score, performed comparably to the IForest model in terms of accuracy and F1 (Minority).

Analysis of Summary Table and Heatmap

The IForest model is recommended for the wbc dataset. Its strengths align well with the dataset's properties, including high skewness, high kurtosis, and structured data. Despite the small sample size, which is a weakness for the IForest model, it still outperformed other models in key metrics.

LLM Recommendations and Justification for Model Selection

Given the high skewness and kurtosis of the wbc dataset, the IForest model is a suitable choice due to its strengths in handling such data. Its superior performance in ROC AUC, F1 (Minority), and Average Precision further justifies this recommendation.

Suggested Preprocessing Steps for wbc

Given the imbalanced nature of the wbc dataset, it is recommended to apply techniques such as SMOTE or ADASYN for oversampling the minority class. Additionally, considering the high skewness and kurtosis, applying transformations such as log or Box-Cox to reduce skewness could be beneficial.

Hyperparameter Tuning Recommendations

For the IForest model, key hyperparameters to consider include the number of estimators and the contamination factor. A grid search or random search could be used to find the optimal values for these hyperparameters.

Final Professional Recommendation

Based on the LLM analysis, the IForest model is recommended for the wbc dataset. This recommendation is supported by the model's high performance in key metrics and its strengths in handling datasets with high skewness and kurtosis. Preprocessing steps such as oversampling and transformations should be considered to further improve model performance.