# Model Evaluation Summary Table for vowels

Model	Symbolic Rank	Symbolic Score	ROC AUC	Accuracy	Average Precision	F1 (Minority)	Precision (Minority)	Recall (Minority)	AUC Rank	Accuracy Rank	F1 Rank
LOF	1	3.8	0.943	0.921	0.3162	0.3575	0.2481	0.64	1	1	2
AutoEncoder	1	3.8	0.9385	0.919	0.5504	0.398	0.2671	0.78	2	2	1
VAE	1	3.8	0.6077	0.8771	0.0536	0.0914	0.0612	0.18	5	5	5
DeepSVDD	1	3.8	0.7634	0.8929	0.2107	0.2041	0.137	0.4	3	3	3
IForest	2	3.2	0.7446	0.8915	0.1171	0.1939	0.1301	0.38	4	4	4

Model Evaluation Heatmap for vowels

vowels - Model Evaluation Heatmap
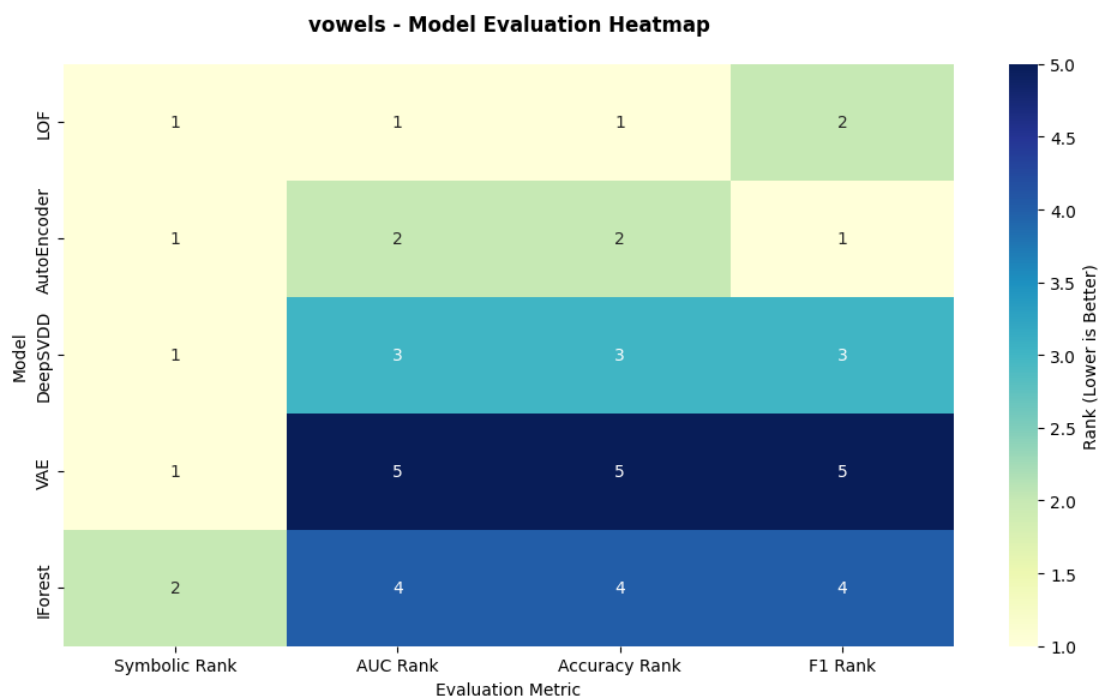
LLM Analysis and Recommendations for vowels

LLM Analysis & Model Recommendation for Vowels Dataset

Overview Summary for Vowels Dataset

The Vowels dataset is a medium-sized, medium-dimensional dataset with 1456 samples and 12 features. It is highly imbalanced with an anomaly ratio of 0.034341. The dataset is clean with no missing values and has low skewness and kurtosis. The data is structured and not sparse.

Comparison of Symbolic vs. Actual Scores

The symbolic scoring system has successfully prioritized high-performing models. The top models according to symbolic scoring are LOF, AutoEncoder, VAE, and DeepSVDD, all with a symbolic score of 3.8. The IForest model has a slightly lower symbolic score of 3.2.

When comparing these symbolic scores with actual evaluation metrics, LOF and AutoEncoder models perform exceptionally well across all metrics. LOF has the highest ROC AUC and Accuracy, while AutoEncoder has the highest Average Precision and F1 (Minority) score. The VAE and DeepSVDD models, despite having high symbolic scores, do not perform as well in actual evaluations.

Model Suitability Summary Table for Vowels

The summary table shows that LOF and AutoEncoder models have the highest overall performance. They rank first and second respectively in most evaluation metrics. The VAE model, despite having a high symbolic score, ranks last in all evaluation metrics. The DeepSVDD and IForest models have intermediate performance.

Model Evaluation Heatmap for Vowels

The heatmap visualization confirms the findings from the summary table. LOF and AutoEncoder models have the highest performance across all metrics, with LOF excelling in ROC AUC and Accuracy, and AutoEncoder in Average Precision and F1 (Minority) score. The VAE model shows low performance across all metrics.

Analysis of Summary Table and Heatmap

The summary table and heatmap provide a comprehensive view of model performance. They confirm that LOF and AutoEncoder models are the top performers. The VAE model, despite its high symbolic score, underperforms in actual evaluations. This discrepancy might be due to the high imbalance in the dataset, which the VAE model might not handle well.

LLM Recommendations and Justification for Model Selection

Based on the analysis, the LOF and AutoEncoder models are recommended for deployment. Both models have high symbolic scores and perform well in actual evaluations. They handle the characteristics of the Vowels dataset well, including its medium size, medium dimensionality, high imbalance, and low skewness and kurtosis.

Suggested Preprocessing Steps for Vowels

Given that the data is clean with no missing values, no preprocessing steps are required for handling missing data. However, considering the high imbalance in the dataset, it might be beneficial to apply oversampling or undersampling techniques to balance the classes.

Hyperparameter Tuning Recommendations

For the LOF model, parameters such as the number of neighbors and the algorithm used for nearest neighbor search can be tuned. For the AutoEncoder model, parameters such as the number of hidden layers, the number of neurons in each layer, and the activation function can be tuned.

Final Professional Recommendation

In conclusion, the LOF and AutoEncoder models are recommended for deployment on the Vowels dataset. These models have shown superior performance in both symbolic scoring and actual evaluations. They are well-suited to handle the dataset's characteristics and are expected to provide reliable anomaly detection.