

The model directory generated by the MLJAR AutoML framework provides a comprehensive, transparent record of the model development process and is publicly available via a GitHub repository (https://github.com/nczub/AhR_Streamlit). This directory is also used directly as the backend for the Streamlit-based web application.

The directory includes subfolders for each machine learning model trained with the available algorithms, as well as files documenting preprocessing, validation, and prediction, and figures visualizing the results.

In addition, the directory includes an Ensemble folder, which contains all files related to the final ensemble model selected by MLJAR. This folder contains configuration files describing the ensemble composition, optimization metric, figures, and final classification threshold, along with a README.md summarizing the ensemble model.

Several files are provided to document key stages of the modeling workflow, including:

- cross-validation settings and results (cv.data),
- feature engineering procedures, including Golden Features (golden_features.json),
- models selected to ensemble model (params.json),
- training progress and execution logs (progress.json),
- dataset (reduced_database.csv),
- general experiment summary (README),
- PNG figures (correlation heatmap, ldb performance boxplot),
- predictions for the ensemble model on the training dataset,
- predictions obtained during k-fold cross-validation,
- predictions for the external validation dataset (in this study, corresponding to the test set).

Together, these files provide full traceability of the machine learning workflow, allowing reproduction of model training, validation, and performance evaluation.