Paper Validation Report for LADS

Correspondence: Yes
Percentages: 85.0%

Conclusion:

The repository code implements multi-agent AutoML workflows for tabular data aligning with the methods described in the paper, including integration of LLM-based code generation and multiple AutoML tools (such as Fedot and LightAutoML). The code supports tabular data ingestion, data splitting, configuration, training, evaluation (including metric extraction), code improvement, and reporting—matching the procedural description in the paper. The benchmark and chat UI facilitates selection and tracking of Kaggle-like tasks. However, while the code provides modules for evaluation metrics, dataset management, and comparison with baselines, explicit scripts or automation for the precise set of 8 Kaggle tasks, along with direct reproduction of all reported normalized performance scores/tables, are not fully surfaced. Some hyper-specific experimental orchestration (e.g., batch running across all datasets) may require additional scripting. Reproduction is feasible but would require manual selection and orchestration. Thus, there is substantial compliance, but minor manual steps or supplemental scripts may be required to exactly match the paper's reported experimental figures.