

Document Validation Report

for AutoTM



Correspondence: Yes

Percentages: 95.0%

Conclusion:

The codebase closely implements the methods, algorithms, and architecture described in the Russian-language documentation. The described components—such as genetic algorithms for hyperparameter optimization, Bayesian optimization with Hyperopt, surrogate modeling, fitness evaluation, and graph/ontology construction based on topic modeling—are all clearly implemented in the provided code. Classes for individuals, populations, parameter mutation/crossover, fitness estimators (including surrogate and distributed modes), as well as dataset preprocessing, topic modeling (with ARTM/BigARTM), LLM-based scoring, and artifact/statistics management are present and align with the technical details in the documentation. The dataset input formats (pandas DataFrame, CSV, etc.), evaluation metrics (coherence, custom metrics, sparsity), and parameter passing structures correspond with the documentation. The main components of the experimental pipeline (preprocessing, model training, optimization, logging) are programmatically reproducible. The only potential minor gap is the absence of explicit scripts or automated workflows for running/reporting consolidated experimental results as described in some output-oriented sections, and slight ambiguity about end-to-end orchestration or UI elements, if any. Overall, nearly all critical system parts are covered for reproducibility and alignment.