How does the scalability of SageMaker Autopilot's feature engineering pipeline affect end-to-end training time

Assignee Research

June 10, 2026

Abstract

AutoML systems provide a black-box solution to machine learning problems by selecting the right way of processing features, choosing an algorithm and tuning the hyperparameters of the entire pipeline. Although these systems perform well on many datasets, there is still a non-negligible number of datasets for which the one-shot solution produced by each particular system would provide sub-par performance. In this paper, we present Amazon SageMaker Autopilot: a fully managed system providing an automated ML solution that can be modified when needed. Given a tabular dataset and the target column

1 Introduction

This paper examines: Amazon SageMaker Autopilot: a white box AutoML solution at scale. Research question: How does the scalability of SageMaker Autopilot's feature engineering pipeline affect end-to-end training time and model accuracy when applied to larger tabular datasets like those in the OpenML-CC18 benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.5/10.

3 Results

13 papers retrieved. 7 claims extracted; 0 independently verified. Quality review score: 3.5/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The simple ϵ-greedy algorithm works better than other more complicated bandit algorithms, such as EXP3 and Rotting bandi	×	0.01
The system has suggested 5 HPs for every pipeline before moving to ϵ-greedy.	×	0.01
One pipeline has finished 5 HP evaluations before moving to ϵ-greedy.	×	0.01
The algorithm can successfully identify the best pipeline on more than 50% of datasets evaluated.	×	0.02
For around 80% of datasets, the algorithm's choice is among the top 3 pipelines.	×	0.05
Using learned 5 zero-shot HP configurations instead of random HPs increases the probability of committing to the best pi	×	0.03
The meta-model is trained with a train collection of datasets and evaluated on a separate test collection of datasets.	×	0.07

References

- http://arxiv.org/abs/2012.08483v2
- http://arxiv.org/abs/2507.05904v1
- http://arxiv.org/abs/2009.02557v1