

Impact of Dataset Size on Feature Selection Consistency and Accuracy of SageMaker Autopilot Versus H2O.ai and TPOT

Assignee Research

June 12, 2026

Abstract

Efforts to make machine learning more widely accessible have led to a rapid increase in Auto-ML tools that aim to automate the process of training and deploying machine learning. To understand how Auto-ML tools are used in practice today, we performed a qualitative study with participants ranging from novice hobbyists to industry researchers who use Auto-ML tools. We present insights into the benefits and deficiencies of existing tools, as well as the respective roles of the human and automation in ML workflows. Finally, we discuss design implications for the future of Auto-ML tool development

1 Introduction

This paper examines: Whither AutoML? Understanding the Role of Automation in Machine Learning Workflows. Research question: What is the impact of dataset size on the feature selection consistency and accuracy of SageMaker Autopilot compared to other AutoML systems like H2O.ai and TPOT when evaluated across multiple domains?.

2 Methodology

Systematic literature search across multiple databases yielded 5 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.3/10.

3 Results

5 papers retrieved. 6 claims extracted; 6 independently verified. Quality review score: 8.3/10.

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
Efforts to make machine learning more widely accessible have led to a rapid increase in Auto-ML tools that aim to automa	✓	0.47
A qualitative study was performed with participants ranging from novice hobbyists to industry researchers who use Auto-M	✓	0.35
The study presents insights into the benefits and deficiencies of existing Auto-ML tools.	✓	0.25
The study discusses the respective roles of the human and automation in ML workflows.	✓	0.23
The paper argues that instead of full automation being the ultimate goal of Auto-ML, designers of these tools should foc	✓	0.41
A range of Auto-ML tools will need to be developed to support varying user goals such as simplicity, reproducibility, an	✓	0.41

References

- <https://doi.org/10.1613/jair.1.11854>
- <https://doi.org/10.1145/3411764.3445306>
- <https://doi.org/10.1007/s10462-024-10726-1>