

CausalMixFT Scaling Laws in Tabular Foundation Models Under Distribution Shift

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: Does the effectiveness of CausalMixFT scale with the size of the tabular foundation model, as measured by accuracy improvements on SuperTab’s distribution shift benchmarks. 8 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Benchmarking Distribution Shift in Tabular Data with TableShift. Research question: Does the effectiveness of CausalMixFT scale with the size of the tabular foundation model, as measured by accuracy improvements on SuperTab’s distribution shift benchmarks?.

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 4.2/10.

3 Results

13 papers retrieved. 8 claims extracted; 0 independently verified. Quality review score: 4.2/10.

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
Food insecurity affects more than 10% of households (13.5 million) across the United States in 2021.	×	0.01
The American Community Survey (ACS) is used as the data source for the Food Stamps task.	×	0.03
The Food Stamps task filters data for low-income adults aged 18-62 in households with at least one child.	×	0.02
An income threshold of \$30000 is used based on the U.S. poverty threshold for a family with one child.	×	0.02
The ACS includes 10 regions: Puerto Rico; New England (Northeast region); Middle Atlantic (Northeast region); East North	×	0.00
East South Central (South region) is used as the holdout domain for the Food Stamps task.	×	0.02
The TableShift benchmark includes tasks such as ASSISTments Next Answer Correct, College Scorecard Low Degree Completion	×	0.04
The TableShift benchmark evaluates models including XGBoost, LightGBM, MLP, FT-Transformer, ResNet, SAINT, NODE, TabTran	×	0.04

References

- <http://arxiv.org/abs/2506.16791v4>
- <http://arxiv.org/abs/2512.03307v1>
- <http://arxiv.org/abs/2312.07577v3>