Novel Metrics for Robust Generative Model Evaluation on Mixed-Type Tabular Data

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: How do the novel metrics proposed for evaluating generative models on tabular data with mixed data types compare in robustness against class imbalance compared to traditional metrics like AUROC, and. 11 claims were extracted from source literature; 3 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Evaluating Generative Models for Tabular Data: Novel Metrics and Benchmarking. Research question: How do the novel metrics proposed for evaluating generative models on tabular data with mixed data types compare in robustness against class imbalance compared to traditional metrics like AUROC, and what is the impact on alignment in low-resource settings?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.5/10.

3 Results

11 papers retrieved. 11 claims extracted; 3 independently verified. Quality review score: 5.5/10.

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
FAED effectively captures generative modeling issues overlooked by existing metrics.	✓	0.28
FPCAD exhibits promising performance but requires further refinements to enhance its reliability.	✓	0.18
FAED successfully detects all synthesized problems in the experimental results on three standard network intrusion detec	×	0.14
FPCAD shows promising but improvable performance in detecting synthesized problems.	×	0.04
Existing metrics fail to identify key issues in generative modeling for tabular data.	×	0.15
FAED and FPCAD are novel and robust metrics for evaluating generative models in the tabular data domain.	✓	0.29
Quality Decrease, Mode Drop, and Mode Collapse issues were embedded into the datasets to simulate real-world generative	×	0.06
FAED, FPCAD, and RFIS were compared with existing metrics in an in-depth comparative analysis.	×	0.14
TSTR is useful for detecting cases where synthetic data only partially represents real data.	×	0.05
TRTS assesses whether synthetic samples introduce patterns absent in real data.	×	0.04
TSTR and TRTS together provide a comprehensive view of synthetic data quality.	×	0.07

References

- http://arxiv.org/abs/2306.13929v1

- http://arxiv.org/abs/2504.20900v1
- http://arxiv.org/abs/2502.17119v2