Diffusion Models vs. VAEs in Large-Scale Tabular Data Generation: Scaling and Efficiency Benchmarks

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How do diffusion models scale in terms of training time and memory efficiency compared to VAEs when applied to large-scale tabular datasets with 100+ features, as benchmarked on the Kaggle Titanic. 19 claims were extracted from source literature; 1 was 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: Diffusion and Flow Matching Models for Tabular Data: A Survey. Research question: How do diffusion models scale in terms of training time and memory efficiency compared to VAEs when applied to large-scale tabular datasets with 100+ features, as benchmarked on the Kaggle Titanic dataset?.

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. 19 claims extracted; 1 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
Formal differential privacy has begun to appear in tabular diffusion models.	×	0.08
Most tabular diffusion works rely on empirical privacy metrics such as distance-to-closest-record, attribute inference,	×	0.06
Recent memorization studies indicate that a small subset of records may dominate memorized generations.	×	0.02
Removing memorization-prone or rare records can change minority-group coverage.	×	0.01
Many diffusion and flow matching models operate as black-box generative systems.	✓	0.17
The anomaly detection method TCCM shows that feature-level residuals can support interpretability.	×	0.04
Explanation tools for synthesis and imputation are underdeveloped compared to anomaly detection methods.	×	0.06
Scaling, encoding, and inverse transformations can change how feature-level evidence is interpreted.	×	0.02
Healthcare models such as FlexGen-EHR and PatientFlow point toward multimodal and longitudinal tabular generation.	×	0.05
Most current benchmarks for tabular data generation remain single-table and static.	×	0.10
Several recent methods combine diffusion or flow matching with autoencoders, transformers, tree models, or feature-token	×	0.13
Diffusion SOS (2022) is a synthesis model for single, generic tables using SDEs.	×	0.03
STaSy (2023) is a synthesis model for single, generic tables using SDEs.	×	0.02
TabDDPM (2023) is a synthesis model for single, generic tables using DDPM+MLD.	×	0.01
CoDi (2023) is a synthesis model for single, generic tables using DDPM+MLD.	×	0.01
AutoDiff (2023) is a synthesis model for single, generic tables supporting any probability path.	×	0.04
MissDiff (2023) is a synthesis model for single, generic tables using SDEs.	×	0.02
Data augmentation for tabular data can be divided into data synthesis and over-sampling.	×	0.09
Over-sampling can be considered a special case of single table synthesis where only a part of the table is generated.	×	0.02

References

- http://arxiv.org/abs/2405.03150v2
- http://arxiv.org/abs/2504.20900v1
- http://arxiv.org/abs/2502.17119v2