

How does flow-matching based synthetic data generation compare to CTGAN in improving minority class recall for

Assignee Research

June 10, 2026

Abstract

Handling imbalance in class distribution when building a classifier over tabular data has been a problem of long-standing interest. One popular approach is augmenting the training dataset with synthetically generated data. While classical augmentation techniques were limited to linear interpolation of existing minority class examples, recently higher capacity deep generative models are providing greater promise. However, handling of imbalance in class distribution when building a deep generative model is also a challenging problem, that has not been studied as extensively as imbalanced class

1 Introduction

This paper examines: Synthetic Tabular Data Generation for Imbalanced Classification: The Surprising Effectiveness of an Overlap Class. Research question: How does flow-matching based synthetic data generation compare to CTGAN in improving minority class recall for imbalanced tabular datasets?.

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 3.3/10.

3 Results

13 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.3/10.

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.

References

- <http://arxiv.org/abs/2412.15657v2>
- <http://arxiv.org/abs/2502.17119v2>
- <http://arxiv.org/abs/2605.14067v1>