

## **Generative AI: A Comparison of Conditional Tabular Generative Adversarial Networks and Conditional Tabular Generative Adversarial Networks with Gaussian Copula in Generating Synthetic Data with Synthetic Data Vault**

**Authors :** Lakshmi Prayaga, Chandra Prayaga. Aaron Wade, Gopi Shankar Mallu, Harsha Satya Pola

**Abstract :** Synthetic data generated by Generative Adversarial Networks and Autoencoders is becoming more common to combat the problem of insufficient data for research purposes. However, generating synthetic data is a tedious task requiring extensive mathematical and programming background. Open-source platforms such as the Synthetic Data Vault (SDV) and Mostly AI have offered a platform that is user-friendly and accessible to non-technical professionals to generate synthetic data to augment existing data for further analysis. The SDV also provides for additions to the generic GAN, such as the Gaussian copula. We present the results from two synthetic data sets (CTGAN data and CTGAN with Gaussian Copula) generated by the SDV and report the findings. The results indicate that the ROC and AUC curves for the data generated by adding the layer of Gaussian copula are much higher than the data generated by the CTGAN.

**Keywords :** synthetic data generation, generative adversarial networks, conditional tabular GAN, Gaussian copula

**Conference Title :** ICMLT 2024 : International Conference on Machine Learning Technologies

**Conference Location :** Stockholm, Sweden

**Conference Dates :** July 15-16, 2024