

# Use Case Brief: Synthetic Data for Real Estate Secured Creditworthiness Validation

## Client Profile and Business Problem

**Client:** Large Regional Private Debt Fund (Alternative Lender).

**Business Problem:** The Fund aims to deploy an advanced Machine Learning (ML) scoring model to quickly and accurately assess the creditworthiness of applicants for small, short-term loans secured by real estate collateral. The goal is to maximize yield while maintaining low default rates in a non-bank lending environment. The Fund needs to rapidly validate the predictive power and stability of its new ML scoring engine across thousands of simulated scenarios before deployment.

## The Challenge: Data Scarcity and Regulatory Friction

The development and validation process is hampered by two critical constraints:

1. **Regulatory Restrictions (GDPR/PII):** Due to the sensitive nature of borrower data (income statements, collateral valuation, and specific repayment history), the Fund cannot easily share its limited historical dataset with third-party model developers or use it in scalable cloud environments for stress testing. This creates a validation bottleneck.
2. **Scarcity of Edge Cases:** The historical dataset lacks sufficient examples of specific high-risk cohorts—for instance, borrowers with variable income streams facing simultaneous minor dips in real estate collateral valuation—which are crucial for robustly testing the model's generalization capabilities at the tail-end of the risk distribution.
3. **Validation Auditability:** Regulators require auditable proof that ML models perform reliably across all potential borrower behaviors, which cannot be achieved using small, non-representative historical samples.

## Northhaven Solution: Custom Generative Model Deployment

Northhaven Analytics provides a dedicated, custom-trained **Synthetic Data Generator (SDG) ML Model** that replicates the Fund's specific lending ecosystem entirely without using a single row of real client data.

### Solution Components:

- **Custom ML Architecture:** The SDG utilizes a **C-CTGAN** framework augmented with a

**Temporal Sequence Model (TSM)** to capture the temporal dependency between income/cash flow and repayment performance.

- **Data Generation:** The model is trained on the Fund's statistical aggregates and schema, allowing it to generate a statistically equivalent synthetic dataset containing millions of unique borrower profiles.

## Key Synthetic Deliverables:

Northhaven generates the core inputs necessary for robust credit scoring model validation:

- **Income Patterns:** Realistic synthetic time-series modeling of fluctuating income and cash flow cycles typical of self-employed borrowers or small businesses.
- **Collateral Profiles:** Correlation-preserving synthetic data linking initial Loan-to-Value (LTV) ratios with simulated changes in **real estate collateral valuation** under various stress conditions.
- **Repayment Logic:** Replication of the behavioral sequence leading up to default or successful repayment, providing rich history for predictive modeling.
- **Early-Warning Indicators (EWI):** Targeted generation of synthetic scenarios showing specific, low-frequency combinations of events (e.g., missed payment coincident with a 15% collateral value drop), essential for EWI model calibration.

## Outcome and Value Proposition

The deployment of the Northhaven SDG model transforms the Fund's risk management capabilities:

- **GDPR-Safe Experimentation:** The Fund gains access to an unlimited, privacy-compliant synthetic data environment, eliminating regulatory barriers for model tuning, stress testing, and external collaboration.
- **Improved Model Generalization:** By generating millions of synthetic, high-fidelity records, including targeted **rare-event scenarios**, the Fund can train and validate its ML scoring model against comprehensive risk profiles, significantly improving model robustness and generalization.
- **Validation of Lending Rules:** The Fund can use the synthetic data to test new internal lending policies (e.g., changes to maximum LTV or income requirements) by simulating their impact before implementation, lowering deployment risk.
- **Quantifiable Fidelity:** The delivered model is accompanied by audit documentation confirming **Synthetic-to-Real Correlation Retention** (Target > 0.95), assuring the Fund and its investors that the synthetic data accurately reflects the real portfolio's risk characteristics.

## Key Takeaway

Northhaven enables the Private Debt Fund to leverage cutting-edge ML models for higher returns and better risk selection by solving the fundamental tension between data privacy,

scalability, and regulatory-mandated validation.