

Northhaven Analytics: Comprehensive Investor Report

Date: December 4, 2025

Prepared For: Venture Capital Due Diligence Committee

Core Value Proposition: Delivery of institution-specific Machine Learning Generative Models (SDGs) for financial data replication, solving the tension between high-scale ML validation and regulatory data privacy (GDPR, SR 11-7).

SECTION 1: Commercial Momentum & Traction Summary (Q4 2025)

Northhaven has successfully moved past the R&D stage, validated a high-margin enterprise delivery model, and established initial commercial traction.

1.1 Early Commercial Engagement & Flagship Proof-of-Concept (PoC)

Northhaven has established a high-value entry point with a critical sector, validating the immediate commercial demand for privacy-compliant risk modeling solutions.

- **First Client Engagement (Late-Stage PoC):** Large Regional Private Debt Fund.
- **Use Case:** Development of a dedicated synthetic data generation model for **creditworthiness checks** concerning small private debt loans secured by real estate. This model is essential for validating the Fund’s internal underwriting models against rare default scenarios.
- **Status:** Late-stage technical discussion completed; model specification phase underway, contingent on final PoC delivery.
- **Initial Contract Value (ICV) Indicator:** The projected margin for the development and deployment of this single custom ML generative engine exceeds **\$200,000 USD** (equivalent to PLN 800,000). This figure validates the high-margin, enterprise nature of our dedicated ML artifact delivery model.

1.2 Market Indicators and Pipeline Overview

Early traction indicators confirm Northhaven’s unique positioning against privacy and scale constraints.

Indicator	Status & Metric	Relevance
Inbound Interest (Fintech/Quant)	Received unsolicited inquiries from five Fintech founders and multiple	Validates broad market fit beyond the Private Debt segment.

	quantitative research teams.	
Technical / Talent Traction	Significant inbound inquiries from software engineers and ML developers expressing interest in employment, engagement, or learning about the project.	Confirms an innovative position within the deep-tech community, securing a strong talent pipeline.
Academic Partnership	Actively finalizing an educational collaboration agreement with a leading University's Quantitative Finance Department.	Establishes early thought leadership and access to advanced research.
Near-Term Pipeline (6-12 Months):	Private Debt, Alternative Lenders, and building relationships with banks by offering free demonstrative datasets.	Focus on securing 2-3 additional high-value custom ML engine development contracts to build portfolio and establish positioning.

SECTION 2: Use Case Deep Dive (Private Debt Fund)

2.1 Business Problem and Regulatory Friction

The Private Debt Fund must deploy an advanced ML scoring model but is blocked by two critical constraints:

1. **Regulatory Restrictions (GDPR/PII):** Sensitive borrower data (income, collateral valuation) cannot be shared with third parties or used in scalable cloud environments for stress testing, creating a validation bottleneck.
2. **Scarcity of Edge Cases:** The historical dataset lacks sufficient examples of high-risk cohorts (e.g., borrowers with variable income streams facing simultaneous minor dips in real estate collateral valuation), compromising the model's **generalization capabilities**.

2.2 Northhaven Solution: Custom Generative Model (SDG)

Northhaven 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.

- **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.
- **Key Synthetic Deliverables:** The model generates high-fidelity synthetic data for **Income Patterns** (time-series modeling of fluctuating income), **Collateral Profiles** (correlation-preserving data linking LTV and simulated collateral valuation changes), and **Early-Warning Indicators (EWI)** (targeted generation of low-frequency default scenarios).

2.3 Outcome and Value Proposition

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

- **GDPR-Safe Experimentation:** Eliminates regulatory barriers, allowing unlimited, privacy-compliant experimentation.
- **Improved Model Generalization:** By generating targeted **rare-event scenarios**, the Fund can validate its ML scoring model against comprehensive risk profiles, significantly improving model robustness.
- **Quantifiable Fidelity:** The delivered model is accompanied by audit documentation confirming **Synthetic-to-Real Correlation Retention** (Target > 0.95), assuring that the synthetic data accurately reflects the real portfolio's risk characteristics.

SECTION 3: Technical Deep Dive & Financial Architecture

3.1 Generative Engine Architecture and Modularity

Our core product is a dedicated ML artifact. The platform is implemented as a cohesive, modular Python library for enterprise integration.

Component	Technical Role	Functionality
Generator (G) / Discriminator (D)	C-CTGAN / TSM Hybrid	Creates synthetic data points; Discriminator reinforces probabilistic constraints derived from financial logic.
model Module	ML Core Artefact	Encapsulates the custom-trained model weights and execution logic.

git_controller	Reproducibility Engine	Manages model artifact versioning using Git/GitHub, ensuring that any synthetic dataset can be precisely reproduced for regulatory auditability (SR 11-7).
-----------------------	------------------------	--

3.2 Performance and Scalability Benchmarks (Market Leader)

The platform is optimized for high-volume financial data generation, demonstrating market-leading sub-linear scaling capabilities.

Output Volume	Generation Time (Approx.)	Schema Complexity
1 Million Rows	approximately 8 minutes	Standard (10 features, 1 sequence)
10 Million Rows	approximately 15 minutes	Standard
1 Billion Rows	approximately 16 hours	High (50 features, multiple sequences)

3.3 Investment Allocation and Runway

The **\$4,000,000 USD** investment ensures a **47-Month Runway** at stabilized Gross Burn, maximizing the potential for achieving technical leadership and commercial validation before the Series A round.

Allocation Area	Value (USD)	Goal
Personnel (12-18 Months)	\$2,000,000	Secure and scale core engineering/quant team (5 FTEs).
R&D / Product Development	\$800,000	Train and certify 4-5 core generative ML models; develop Continuous Learning Loop features.

Sales & Marketing	\$600,000	Target enterprise financial institutions; drive PoC conversions.
Working Capital & Contingency	\$400,000	Maintain liquidity; buffer for legal/cloud spikes.

SECTION 4: Detailed Financial Projections (M1-M12)

This section provides a granular view of the revenue drivers, cost structure, and milestone timeline underpinning the \$4M investment thesis.

4.1 Revenue Sources & Cost Components

Revenue Sources	Cost Components
Custom ML Model Development: High-value, one-off contracts for initial engine delivery (~\$200k each).	Personnel: Hiring 2-3 Software Engineers and 1 Quantitative Analyst.
Custom Synthetic Dataset Projects: Large-volume data generation runs for stress testing/simulations.	Cloud Compute: Training CTGAN/TSM models on high-performance GPUs/TPUs.
Subscription (Post-M12): Recurring revenue for model refresh, maintenance, and advisory services.	Legal & Compliance: Establishing IP, data governance, and regulatory assurance.
	Basic Operational Costs: G&A, software licensing, and office expenses.

4.2 Projected Monthly Operating Costs (Burn Rate)

Category	Months 1-3 (Setup/PoC)	Months 4-12 (Scaling)	Annualized Run Rate (M12)
Personnel (Salary + Benefits)	\$46,000 (4 FTEs)	\$70,000 (5 FTEs)	\$840,000
Cloud Compute (R&D)	\$4,000	\$8,000	\$96,000

Legal & Compliance	\$15,000 (Initial Setup)	\$2,000 (Retainer)	\$29,000
Operational & G&A	\$3,000	\$4,000	\$45,000
Total Gross Burn (Approx.)	\$68,000	\$84,000	\$1,010,000

4.3 Revenue Plan and Milestone Timeline (M1-M12)

Month	Phase & Focus	Key Milestone	Projected Revenue (USD)
M1 - M6	Productization & PoC	Finalize the modular Python library (MVP); complete Private Debt Fund Model Specification.	\$0
M7	First Revenue	Project 1 (Private Debt Fund) ML Engine Deployment & Invoice.	\$200,000
M8 - M10	Pipeline Acceleration	Secure 2nd client contract (Alternative Lender); continue developing 2nd client model.	\$0
M11	Second Revenue	Project 2 (Alternative Lender) ML Engine Deployment & Invoice.	\$200,000
M12	Scaling Revenue	Project 3 (Bank PoC) Conversion & Deployment.	\$200,000

Cumulative Revenue (M12)			\$600,000
--------------------------	--	--	-----------