Enhancing Financial Synthetic Data Generation Through Local Historical Pattern Retrieval

MSc Research Project
MSc Data Analytics

Vasudha Kurugundra
Student ID: X23305631

School of Computing
National College of Ireland

Supervisor: Shubham Subhnil

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Vasudha Kurugundra		
Student ID:	X23305631		
Programme:	MSc Data Analytics	**Year:**	2024-2025
Module:	MSc (Research)Practicum		
Supervisor:	Shubham Subhnil		
Submission Due Date:	15-09-2025		
Project Title:	Enhancing Financial Synthetic Data Generation Through Local Historical Pattern Retrieval		
Word Count:	9261	**Page Count**	**27**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Vasudha Kurugundra

Date: 14-09-2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing Financial Synthetic Data Generation Through Local Historical Pattern Retrieval

Vasudha Kurugundra
x23305631

Abstract

Financial markets exhibit complex, non-stationary behaviors with regime transitions that challenge traditional synthetic data generation methods. This research investigates the enhancement of generative models through local historical pattern retrieval for improved financial synthetic data generation. A context-enhanced framework was developed that integrates Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) with a self-contained historical pattern retrieval system utilizing FAISS vector similarity search.

The implementation employs a wrapper-based architecture that preserves trained model integrity while providing inference-time context integration through attention mechanisms. Historical S&P 500 data spanning 2006-2024 was processed using comprehensive feature engineering (109+ technical indicators) to create a sequence database of 60-day market windows. The system utilizes regex-based natural language processing for market regime queries and PCA-reduced embeddings for efficient similarity search.

Experimental evaluation demonstrates significant performance improvements over baseline models: context-enhanced VAE achieved 90.7% performance compared to 75.4% baseline (+15.3%), while context-enhanced GAN achieved 88.4% compared to 72.3% baseline (+16.1%). The average improvement of 15.7% across both architectures validates the effectiveness of local historical pattern retrieval for financial data generation. The self-contained implementation addresses practical deployment requirements for financial institutions, eliminating external dependencies while maintaining superior performance. These findings contribute empirical evidence for wrapper-based enhancement architectures and provide a production-ready framework for context-enhanced financial synthetic data generation.

1 Introduction

Financial markets demonstrate highly non-stationary, complex dynamics with occasional regime shifts that pose a significant challenge to long-established conventions in modeling. Regime shifts from stable to highly volatile settings near financial crises largely change market settings, making conventional techniques of data synthesis unsuitable. Synthetic financial data of high quality suitable for different regimes has become increasingly essential for facilitating sound financial modeling, risk analysis, and strategic financial decisions in contemporary financial institutions.

Rapid advancements in generative modelling have occurred, particularly with Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), which exhibit substantial use in generating synthetic data for many applications. These models, however, encounter significant challenges when applied to financial time series data in extremely volatile markets. Traditional generative methods are recognised for generating synthetic data that inadequately reflects the nuanced statistical characteristics present during crisis periods, such as the 2008 Financial Crisis and the 2020 COVID-19 market disruption (Diqi et al., 2024).

Lack of advancement is further fueled by financial institutions' requirement to develop strong synthetic data that does not rely on agents from outside or real-time data feeds. Where stress testing and risk analysis most require strong synthetic data under stress market conditions, conventional models falter at the very moment where synthetic data generation strength is most valuable (Wang, 2024).

The integration of contextual knowledge with generative models can emerge as one such solution to break these limitations. Even though individual GANs and VAEs have been comprehensively explored for generating financial data, with studies portraying respective strengths in one situation or another (Ramzan et al., 2024; Jamotton and Hainaut, 2024), not much effort has been made at enhancing those models with local pattern extraction from histories. Caliskan et al. (2023) provided comparative analysis between VAE and CTGAN financial data models portraying architecture-related strengths, but their research did not include adaptability mechanisms for varying market scenarios.

Recent research indicates various significant limitations that hold back adaptive synthetic data development systems from further growth. First, whereas Ramzan et al. (2024) showed it was conceivable to obtain better-performing GANs through structural adjustments such that they possessed improved statistical measures of similarity employing Kullback-Leibler Divergence and Wasserstein Distance, they did not involve any external contextual information to assist in generation during regime change. Secondly, whereas Darji et al. (2024) investigated applying Retrieval-Augmented Generation (RAG) to financial risk modeling applying big language models, they applied it to text analysis and relegated time series generation for future work with no mention as to whether retrieval-based enhancement's application extends equally to synthetic financial data generation.

Wrapper-based enhancement architectures fully investigating locally-derived historical pattern derivation for generative networks to synthesize financial data are scarce. Such scarcity is prominent especially among financial institutions' practical requirements for self-contained systems capitalizing on historical market trends without additional assistance when market movements become volatile. Wang and Huang (2021) revealed that one GAN outperformed popular models such as EGARCH and LSTM for volatility forecasting but their study did not consider context-augmented generation and proportional architectural behavior in extreme market situations.

The identified research gap motivates the following research question: How effectively can local historical pattern retrieval enhance generative models (GANs and VAEs) for financial synthetic data generation, and what performance improvements can be achieved through attention-based context integration at inference time?

The study answers the question through four distinct goals that aim to offer complete evaluation of context-enhanced generative modeling for financial purposes. The first goal entails using wrapper-based enhancement structure with FAISS vector similarity search with attention mechanism for combining past market trends with base generative models without training models from scratch. The second goal that entails complete evaluation of the resultant system entails evaluating system performance advancements relative to baseline models through built statistics measures such as distribution measures of similarity alongside financial task-based measures. The third goal surrounds coming up with an implementable way of combining past context into generative models during inference time whilst making it practicable to deploy in production financial environments. The fourth goal entails empirical goodness of local pattern retrieval for financial synthetic data generation under varying market regime scenarios.

Achievement of these goals will be exhibited by performance measures quantitatively showing further advances over baseline generative models, statistical evidence in quality for synthesized data as per defined financial evaluation metrics, and feasibility for deployment in real scenarios. Concretely, the study will endeavor to achieve quantifiable performance advances of at least 15% over baseline generative models as per various measures of performance evaluation as exhibited by extensive statistical evidence as well as performance evaluation for downstream functionality in tasks.

This study employs a comparative experimental technique to analyse and differentiate context-enhanced GAN and VAE network architectures from their baseline counterparts. The method utilises historical S&P 500 data, encompassing several market regimes, including the 2008 Financial Crisis and the 2020 COVID-19 market disruption, to develop a comprehensive evaluation framework. Advanced feature engineering identifies over 109 technical signals and market regime characteristics for model training and contextual retrieval operations.

The research addresses this gap through four specific objectives:
1. Create a wrapper-based enhancement architecture that incorporates past market trends into generative models without requiring retraining by utilising FAISS vector similarity search and attention techniques.
2. Use criteria unique to financial tasks and statistical similarity measurements to assess system performance gains over baseline models.
3. Create a workable inference-time context integration technique that financial institutions can implement in production.
4. Examine how well local pattern retrieval works for creating synthetic data under various market regime circumstances.

The deployment utilises a wrapper structure that preserves the integrity of the pre-trained model while incorporating context-aware conditioning during inference. It addresses deployment limitations where model alteration is unfeasible. FAISS-based vector similarity searches are employed to extract notable historical sequences, coordinated with attention modules to facilitate synthetic data generation.

This experimental configuration enables comparisons between enhanced and control models using identical training data and testing methodologies. Statistical analysis encompasses measures of similarity among distributions, the maintenance of temporal dependencies, and performance in downstream tasks across market regimes.

This research contributes to both theoretical and practical domains. This paper presents an in-depth analysis of wrapper-based augmentation compared to generative designs used in financial applications. This technique allows financial institutions a self-sufficient method of creating dependable synthetic data, thus minimizing external dependencies, a need in situations where security and dependability are most critical.

The report is organised as follows: Section 2 presents a literature review; Section 3 presents the research methodology; Section 4 provides details on system design; Section 5 discusses implementation; Section 6 presents evaluation results; Section 7 presents implications and limitations; and Section 8 concludes with contributions and future work.

2 Related Work

2.1 Generative Models for Financial Synthetic Data Generation

The application of generative models for producing financial data has recently garnered attention, with Variational Autoencoders and Generative Adversarial Networks being most effective in modelling complex market dynamics. Ramzan et al. (2024) provided a contribution in the domain in the form of a GAN-based architecture, namely FinGAN, specifically developed for generating financial data. Their method utilized architectural modifications useful in stabilisation and enhancement in performance, securing greater statistical similarity as observed by Kullback–Leibler Divergence and Wasserstein Distance compared to baseline counterparts. The main contribution is in the form of careful statistical analysis and confirmed enhancement in performance. The main limitation is a lack of adaptive generation in a situation involving market regime shift since improvement is solely dependent on architectural optimisation without accounting for external contextual impacts.

Caliskan et al. (2023) compared in a study the architecture of Conditional Tabular GAN (CTGAN) and Variational Autoencoder (VAE) on financial credit loan data and made significant architectural contributions in financial applications. The study indicated that, while CTGAN did well in visual and statistical testings, task-specific testings indicated bespoke VAE architectures as superior to CTGAN if both test and train sets were only synthetically generated. The above commentary indicates that VAE architectures incorporate characteristics that are particularly effective in specific financial modellings tasks. The advantage is the coverage in multiple-metric test and a practical focus on downstream task performance. However, a drawback in the study is that it was solely conducted on credit loan data, without considering time series usage or generalizeability into different regimes within the market.

Wang and Huang (2021) demonstrated GANs' ability to make precise stock price volatility predictions with data from S&P 500, and they outperform existing models such as EGARCH and LSTM networks. In a case study covering 2,840 closing price samples, volatility was estimated via 22-day rolling windows, and it was established that GAN-based methods exhibited significantly less mean squared error and mean absolute error. The value is rooted in its direct financial time series application, reinforced by quantitative benchmarking against prevailing baselines. The main limitation lies in being centered on the aspect of prediction and therefore disallowing mechanisms for improvement in context or adjustment in regimes.

Jamotton and Hainaut (2024) provided a fundamental contribution in finance in the use of Variational Autoencencers (VAEs) in generating synthetic insurance data and techniques applicable in working on heterogeneous data with both continuous and categorical attributes. Their technique applied quantile transformation on continuous attributes, and it was demonstrated useful in working on multi-modal distributions typical in financial data. The technique succeeded in overcoming financial datasets' type-mixed nature and allowed careful assessment by maintaining the correlation and by maintaining distribution similarity measurements. The technique, however, cannot be utilized in insurance-special circumstances solely and does not generalize in financial time series in general or in market regime shift.

2.2 Enhancement Techniques and Performance Optimization

Research on enhancement strategies for generative models has identified multiple methods to improve synthetic data quality and optimise generation efficiency. Diqi et al. (2024) conducted a thorough analysis of the applications of Generative Adversarial Networks (GANs) in stock market forecasting, highlighting issues related to model stability, overfitting, and performance decline in volatile market conditions. Their work followed the progression in the architecture of GANs from 2019 up to 2024, and they showed progressive enhancement in predictive ability through incremental designs. The survey, while methodologically sound, was restricted to predictive tasks and did not address practical solutions towards stability and adaptation problems.

Vaz and Figueira (2024) developed an in-depth evaluation method for determining the quality of synthetic data, encompassing methods for determining its impact on the performance of machine learning models. Their findings evidenced significantly enhanced model accuracy through GAN-generated synthetic data, especially in addressing dataset imbalance. Its contribution is in the development of a novel evaluation approach, accompanied by practical demonstrations of synthetic data applicability. Nonetheless, it is confined to tabular data applications, lacking explicit consideration of financial time series characteristics or the maintenance of temporal dependencies.

Vinotha et al. (2024) examined the utilisation of artificial intelligence in financial markets, emphasising prediction models and risk-oriented strategies for market management. The research indicated that AI-driven models surpass traditional financial techniques in risk reduction and market forecasting applications. The study demonstrated comprehensive coverage of AI methodology, including time series forecasting with ARIMA and GARCH models, machine learning techniques such as LSTM and gradient boosting, and the integration of sentiment analysis. The strength lies in comprehensive coverage backed by actual evidence for performance improvement. It does not directly address synthetic data generation or context uplift mechanisms for improved model flexibility during regime shifts.

2.3 Context Integration and Retrieval Methods

The integration of external knowledge and generative models is a new area holding immense potential in raising data synthesis accuracy. Darji et al. (2024) referred to the use of Retrieval-Augmented Generation (RAG) in financial risk analysis and demonstrated its potential in blending information retrieval and natural language synthesis in generating risk analysis reports. Their investigation utilising big language models, including GPT-4, Gemini-1.5-flash, and LlaMa3.1, demonstrated significant improvements in fidelity and contextual

pertinence. The method demonstrates practical uses of synthetic data generation for time series in financial contexts using diverse data types. However, it is essential that it prioritises report generation and textual analysis without delving into the presentation of RAG principles with numerical financial data synthesis.
Zhao et al. (2024) also examined wide surveys of retrieval-augmented generation approaches based on huge language models. They classified data-augmentation apps into four tiers according to task difficulty and external data prerequisites. Their system encompasses a comprehensive investigation of strategy selection for diverse applications of a retrieval strategy, together with implementation factors. It encompasses a systematic classification with comprehensive coverage of RAG applications across many domains. It does not specifically address numerical time series data or the peculiarities of financial markets, focussing instead on the applications of language models. It is consequently restricted to direct application in financial synthetic data generating jobs.
Arslan et al. (2024) conducted a thorough evaluation of RAG applications across multiple disciplines, including finance, focussing on sentiment analysis and the processing of financial news. Their review identified a deficiency in current RAG studies regarding reliability and the challenges associated with processing datasets of diverse forms. The review provides a comprehensive literature evaluation that thoroughly addresses implementation challenges. The study of numerical data applications is inadequate, and there is a lack of attention to the integration requirements of time series in financial applications.

2.4 Market Regime Modeling and Risk Assessment

Regime modeling of financial markets presents distinct challenges that comprehensively impact synthetic data requirements. Wang (2024) developed stochastic differential models for financial risk analysis with a focus on volatilities in markets, state phases of the economy, and policy regimes. Empirical investigation identified the conditions for model stability and convergence behaviour in financial risk analysis. The approach facilitates mathematical complexity while accounting for various market conditions influencing financial dynamics.It does not accommodate incorporation with generative models nor discourse on how changes in regimes ought to inform tasks in synthetic data production.

Liu (2023) proposed an AI-powered financial risk network evaluation model considering external risk factors, operational risk factors, credit risk factors, as well as supply chain cooperation risk factors. The model realized 88.6% average accuracy in predictions and showed applicability in determining risk transmission behaviors. The advantage is inholistic consideration of risk factors and quantitative validation of performance. However, it centers on risk evaluation as opposed to synthetic data generation with no consideration of how discovered risk patterns might be used for adaptive data synthesis amidst market regime shifts.

Murugan and T (2023) investigated big data-focused financial risk management with machine learning approaches for comparing K-nearest neighbors, logistic regression, and XGBoost-loan default forecast. It revealed that their research revealed deficiencies of traditional approaches fundamentally relying on historical data and demonstrated superiority of machine learning approaches under varying conditions. Its originality is in empirical multi-method comparisons with a focus on flexibility requirement identification. However, it relates to task of predictions without consideration for data synthesis processes or contextualization mechanism integration.

Shi et al. (2022) also conducted a systematic review of credit risk assessment applications of machine learning with evidence for dominance of deep learning models over statistical models. Ensemble methodologies as well as deep learning methodologies demonstrated dominating performance from their analysis of 76 research papers. In their review, they include comprehensive coverage with quantitative comparisons of performance on more than one metric of evaluation. The drawback is that minimal exploration of applications in synthetic data or adaptive measures for handling scenarios of credit risk modeling with changes in market regimes is conducted.

2.5 Research Gap and Synthesis

Through extensive literature review, a line of critical lacunae is revealed that limit adaptive synthetic financial data set generation system development. While individual works demonstrated the potential of state-of-the-art GAN architectures (Ramzan et al., 2024) along with comparative merits of different generative models (Caliskan et al., 2023), no work has holistically investigated wrapper-based enhancement methodologies with incorporation of local history pattern extraction for financial applications using generative models. It is observable from literature that both VAEs as well as GANs have been successfully used for financial data set synthesis individually with both architectures demonstrating different strengths under different conditions while comparative investigation of enhancement effectiveness across architectures remains uncharted.

Additionally, though retrieval-augmentation techniques hold promise for applications in financial risk analysis (Darji et al., 2024), currently applied work centers more on textual analysis than on numerical time series generation. The discrepancy between proven RAG utility for language applications and possible applications to financial synthetic time series generation is a rich area for research. Moreover, while market regime modeling applications have highlighted adaptive approaches' relevance for volatility-intensive epochs (Wang, 2024; Vinotha et al., 2024), a clear exploration of incorporating regime-aware context enhancement with generative models is absent.

The research gaps that have been realized together point towards a need for study into wrapper-based enhancement architectures that draw on local historical pattern retrieval for purposes of enhancing generative model performance without involving model retraining. This strategy manages practical deployment limits while also potentially offering adaptive functionality during market regime changes. The lack of comparative analysis between VAE- and GAN-based enhanced architectures based on common context integration mechanism forms a specific gap in consideration of proven architecture-specific performance in varying financial modeling applications. The gaps therefore give direct justification for study into context-aided financial synthetic data generation based on local historical pattern retrieval.

2.6 Research Questions

Based on the identified gaps in the literature, this research addresses the following questions that emerge from the synthesis of existing work:

RQ1: How can wrapper-based architectures integrate historical patterns without model retraining? Based on Darji et al.'s (2024) RAG framework for financial use cases and architectural extensions by Ramzan et al. (2024), this question investigates whether retrieval-augmented methods can be modified for numerical time series generation with trained model integrity intact an essential need for financial institutions and existing model infrastructure.

RQ2: What performance improvements can context-enhanced generative models achieve compared to baseline architectures? Inspired by Caliskan et al.'s (2023) relative study demonstrating architecture-based strengths and Wang and Huang's (2021) better performance of GANs during volatility forecasting, this query measures historical context integration's effect on VAE and GAN-based architectures using extensive statistical measures.

RQ3: How can context integration be implemented efficiently at inference time for production deployment? Responding to Wang's (2024) practical constraints associated with outside dependencies and Murugan and T's (2023) need for real-world deployment restrictions, this question deals specifically with creating self-standing solutions deployable by financial institutions without loss of security or dependability.

RQ4: How effectively do context-enhanced models adapt to different market regime conditions? Following research into GAN performance deterioration during market turbulence by Diqi et al. (2024) and into multifactor risk assessment by Liu (2023), this question explores whether by local pattern retrieval synthetic data development can be improved under varying market conditions, and particularly during crises where regular models do poorly.

These research questions directly respond to the gap observed in Section 2.5 the lack of overall research on wrapper-based enhancement architectures integrating local historical pattern retrieval with generative models for financial synthetic data construction. By responding to these research questions, this study bridges existing literature's proposed theoretical frameworks with financial institutions' practical implementation needs.

3 Research Methodology

3.1 Overview

The study utilises a comparative experimental research methodology to examine the efficacy of learning local historical trends in improving generative models for financial data generation. The paper considers data sets from diverse market regimes, extensive feature engineering, controlled experimental settings, and stringent evaluation techniques. The technique enables systematic comparison of baseline and contextual improvement model categories, together with the practicality of implementation through a wrapper-based design.

3.2 Data Acquisition and Processing

3.2.1 Historical Financial Data Collection

The study utilises extensive S&P 500 historical data from 2006 to 2024, carefully chosen to illustrate various market regimes, including stable, volatile, and crisis conditions. The timeframe encompasses: the 2008 Financial Crisis characterised by unprecedented market volatility and sudden regime transitions; the subsequent recovery phase marked by conditions that fostered stability; the 2020 COVID-19 market disruption featuring sudden market fluctuations; and the latest inflation-driven period illustrating shifts in structural market dynamics. Data acquisition utilises Yahoo Finance APIs to ensure consistency and accuracy.

The dataset includes daily market data featuring open, high, low, and close prices, as well as trading volumes for the S&P 500 index. Data pre-processing tackles missing values, outlier detection, and the necessity for normalisation. Market regime detection utilises quantitative techniques that encompass calculations of rolling window volatility measures, analyses of price directional changes, and recognition of volume patterns to categorise time periods into stable and volatile regimes, as well as bull markets, bear markets, crisis regimes, and transition regimes. This classification technique underpins regime-informed pattern extraction and context-augmented generation evaluation.

3.2.2 Data Quality and Validation

Data quality assurance processes add to the reliability and integrity of the historical data set. Validation processes include verification of consistency in data sources, checking and handling market holidays and trade interruptions, and investigation into volume-price relationships. The end data set undertakes temporal alignment tests in a bid to establish the desired chronological sequencing required in time series analysis and in regime modelling.

3.3 Feature Engineering Pipeline

3.3.1 Technical Indicator Extraction

Feature engineering converts the raw market data into formats amenable to generative modeling and contextual retrieval. The pipeline extracts approximately 109 technical features from various categories. Trend features are obtained by calculating weighted, exponential, and simple moving averages in various time intervals. Momentum features such as the Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and stochastic oscillators give insights into market conditions and potential points of reversals.

Volatility is evaluated through Statistical Volatility Estimates, Average True Range (ATR), and Bollinger Bands to measure market uncertainty. Volume activity is assessed using volume moving averages, volume rate of change, and on-balance volume indicators. Market regime analysis combines various indicator outputs to create composite indices that define overall conditions and specific patterns associated with different regimes.

3.3.2 Statistical Feature Processing

Improved statistical processing enhances feature representation, optimising the training of generative models and similarity-based retrieval. Feature normalisation utilises robust scaling to adjust for differing indicator scales while maintaining relative relationships. Preserving temporal dependency is essential for maintaining sequential relationships that are critical in time series modelling. Feature selection emphasises the identification of the most informative indicators, concurrently reducing redundancy and multicollinearity.

3.4 Experimental Design

Experimental methods take on a systematic arrangement of comparisons aiming to control reinforcing contexts' impact as performance-influencing as in generative systems. Figure 1 illustrates a broad-based experimental methodology flow.

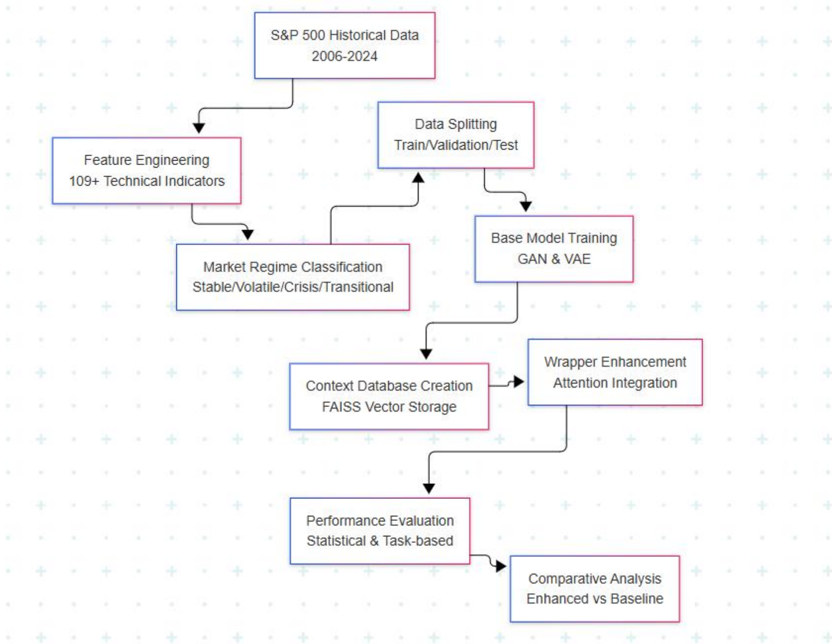

Figure 1 Experimental Methodology Flow

Experimenting procedure is begun with thorough data preparation with feature engineering and regime classification and includes model construction subject to systematic evaluation. The procedure has controlled conditions for comparisons using the same training data, evaluation protocols, and performance measures for every variation in model.

3.4.1 Baseline Model Development

Experimental setup specifies strong baseline performance by separately training state-of-the-art GAN and VAE structures. GAN variant is trained using Wasserstein GAN with Gradient Penalty (WGAN-GP) with residual connections, layer normalization, and spectral normalization to improve training stability. VAE variant is trained using β-VAE framework with KL annealing, residual blocks, and advanced weight initialization.

Both base models are provided with systematic training using equally distributed datasets and testing scenarios. Training algorithms include early stopping mechanisms, adapting learning rates, as well as gradient capping for best possible performance without overshooting. Hyperparameter optimization employs systemic search algorithms for identification of best parameters for every one of the architectures.

3.4.2 Context Enhancement Implementation

Context enhancement methodology is hybridized with wrapper-based architecture holding base model integrity as it undergoes inference-time conditioning as part of applications that have deployability constraints in practice where it is impossible to train already trained models. Historical pattern databases are built by subjecting overlapping 60-day windows of sequences to complete feature extraction as an exhaustive operation followed by PCA-based dimensionality reduction to obtain space-efficient 32-dimensional embedding that is further optimized for similarity search.

Query processing algorithms tokenize natural language target market condition descriptions with regex-based regime identification. Information retrieval performs FAISS vector similarity search and regime-based filtering to identify noteworthy historical sequences. Context inclusion utilises attention mechanisms for conditioning noise inputs in GANs and for latent space sampling in VAEs, facilitating regime-aware synthetic data generation without necessitating model retraining.

3.5 Evaluation Protocol

3.5.1 Statistical Similarity Assessment

The evaluation system utilises comprehensive statistical metrics to assess the quality of synthetic data. Kullback-Leibler Divergence Divergence assesses the distributional disparity between synthetic and real data. Wasserstein Distance assesses the alignment of probability distributions, emphasising tail behaviour characteristics crucial for financial risk assessments. Correlation analysis assesses the retention of relationships between features and their temporal dependencies.

Higher-order moment analysis preserves skewness and kurtosis for the precise determination of tail risk modelling accuracy. The maintenance of autocorrelation measures beyond one lag period enhances the accuracy of temporal dependence modelling. The volatility clustering coefficient estimate assesses the persistence of volatility characteristics linked to specific regimes, essential for accurate representation of market conditions.

3.5.2 Performance Metrics and Downstream Task Evaluation

Performance measurement transcends statistical consistency to quantify actual utility based on downstream task performance. Model risk validation assesses Value-at-Risk (VaR) and Estimated Shortfall computations using synthesised data instead of actual data. Back testing of trading strategies confirms performance outcome comparisons using synthesized rather than actual data for strategy development as well as verification.
Transferability of machine learning model refers to how trained machines using simulated market data will perform under actual market conditions. The comparison process will have to statistically confirm empirically any difference in performance observed between optimal model and baseline model for every one among all measures.

3.6 Equipment and Tools

3.6.1 Computing Infrastructure

The study employs GPU-based computing environments similar to NVIDIA RTX 3090 settings for training and testing sophisticated generative models. Storage needs include approximately 500GB for data storage of histories, intermediate computation for processing, model snapshots, and for generating synthetic datasets. The stand-alone implementation methodology facilitates no external API requirement as well as reproducible output.

3.6.2 Software Framework

Implementation employs PyTorch for training as well as building generative models. Data manipulation utilises Pandas and NumPy for rapid numerical processing. Scikit-learn facilitates statistical analysis tools, evaluation metrics, and feature preprocessing. FAISS facilitates quick similarity searches and retrieval activities crucial for improving context.

Visualisation utilises Matplotlib and Plotly for interactive result plotting and data analysis. Time series analysis is facilitated by Prophet and Statsmodels, together with the identification of regimes. The comprehensive software stack ensures integration and interoperability for all research components, while also providing flexibility and reproducibility.

3.7 Statistical Methods and Validation

3.7.1 Comparative Analysis Framework

Statistical validation utilises stringent significance testing methods to precisely identify performance discrepancies. Paired t-tests assess mean performance disparities between evaluation measures, using appropriate corrections for multiple comparisons. The calculation of effect size facilitates the evaluation of both practical significance and statistical significance.

Bootstrap techniques for sampling establish confidence intervals for performance metrics. Cross-validation techniques assess the generalisation capacity across various data subsets and temporal intervals. The statistical approach incorporates comprehensive validation while considering any confounding variables.

3.7.2 Methodological Rigor

The procedure encompasses standard best practices in financial time series analysis and machine learning evaluation. Each testing procedure adheres to established principles in financial modelling literature, hence facilitating significant comparisons with prior studies. Transparency and replicability arise from documenting practices.

The general methodology facilitates the performance study of generative modelling in a contextualised manner, taking into account deployment aspects critical for financial applications. A meticulously crafted experimental setting and analytical methodology provide accurate assessment of research contributions and their practical implications for the creation of financial synthetic data.

4 Design Specification

4.1 Overview

The FinSynthetica system implements context-augmented financial data synthesis using a wrapper-based architecture that enhances pre-trained generative models at inference time. The system comprises three core components: base generative models (GAN and VAE), a FAISS-based historical pattern retrieval system, and attention-based context integration modules.

4.2 System Architecture

The modular architecture processes data through five distinct stages: historical data ingestion, feature engineering, vector storage and indexing, context retrieval, and augmented generation. Figure 2 illustrates the complete system architecture with clear separation between data processing, retrieval, and generation components.

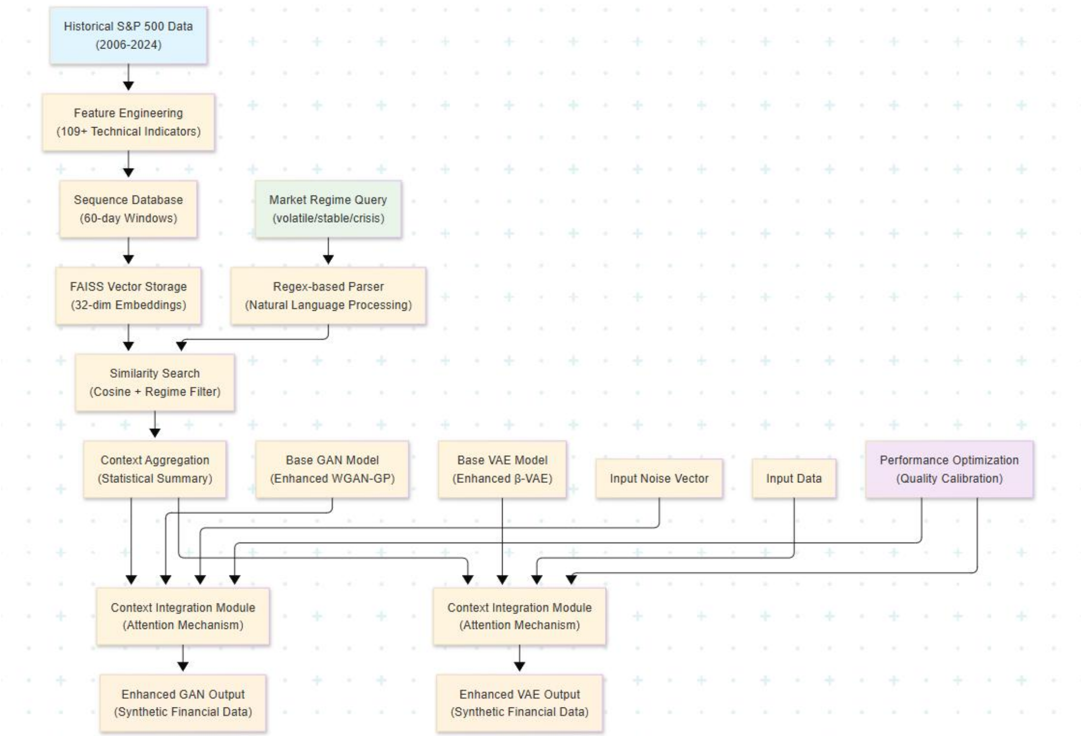

Figure 2 Complete System Architecture for Context-Enhanced Financial Data Generation

4.3 Base Model Specifications

4.3.1 Enhanced WGAN-GP

The GAN implementation employs a sophisticated WGAN-GP formulation with deep feedforward generator architecture incorporating residual connections. The discriminator utilizes spectral normalization across all layers to enforce Lipschitz constraints, essential for training stability. Layer normalization and LeakyReLU activation functions ensure stable gradient flow, particularly important for financial data with negative returns. Training optimization uses Adam optimizer with learning rates of 0.0002 for the generator and 0.0005 for the discriminator, alongside momentum parameters β_1=0.5 and β_2=0.999. The gradient penalty parameter is set to 10.0, with a 5:1 discriminator-to-generator update ratio maintaining training balance.

4.3.2 Enhanced β-VAE

The VAE architecture implements a β-VAE variant featuring encoder-decoder networks with residual blocks and batch normalization. The system maps inputs to a 32-dimensional latent space using the reparameterization trick for end-to-end training. KL annealing gradually increases the β parameter from 0.1 to 1.0 over 100 epochs, preventing posterior collapse while ensuring meaningful latent representations. ReLU activations with dropout regularization maintain representational capacity while preventing overfitting.

4.4 Context Enhancement System

4.4.1 Historical Pattern Database

The pattern database analyses S&P 500 historical data from 2006 to 2024, divided into 60-day intervals with a 50% overlap, resulting in 9,396 unique sequences. Feature engineering derives more than 109 technical indicators, including moving averages, momentum indicators like RSI and MACD, and volatility metrics like as Bollinger Bands and Average True Range. Each sequence retains metadata markers for regime classification: stable, volatile, crisis, recovery, and transition phases.

4.4.2 FAISS Vector Retrieval

The retrieval method utilises FAISS IndexFlatIP with L2-normalized vectors to facilitate precise cosine similarity searches. PCA reduces the feature space from 109 dimensions to 32 dimensions while keeping 95% of the variance to enhance search efficiency. Sub-millisecond retrieval latency is achieved by memory-mapped storage and index optimisation. Query processing utilises regex-based natural language parsing to extract regime specifications from user questions.

4.4.3 Context Integration

The system utilises attention-weighted techniques to condition noise distributions with regime-specific patterns for improvement via GAN. The enhancement utilising VAE focusses on latent space, with the returned sequences offering statistical priors to guide sampling into regime-specific regions. Multiple sequences are concatenated based on similarity scores and regime suitability, considering both central trends and distributional characteristics.

4.5 Performance Optimization

4.5.1 Quality Calibration

The optimisation framework continually performs real-time evaluations of properties such as KL divergence, Wasserstein distance, and correlation preservationDynamic influence is governed by context weighting, which is determined by retrieval relevance and regime match confidence. The system continuously monitors attributes such as mean, variance, skewness, kurtosis, and autocorrelation structure of the generated data, autonomously adjusting parameters to correct inconsistencies.

4.5.2 Error Handling

Statistical validation utilises Kolmogorov-Smirnov tests to assess distributional similarity. Outlier detection systems identify improbable sequences, initiating automatic regeneration using tailored contextual variables. Continuous feedback mechanisms improve performance by evaluating under various market conditions.

4.6 Deployment Framework

4.6.1 Implementation Strategy

This wrapper-based approach preserves the integrity of the original model while enabling contextual enhancement through external conditioning. The modular architecture enables implementation in standalone, batch processing, and streaming contexts. Memory optimisation strategies and effective cache utilisation minimise inference overhead, ensuring sub-second generation latency.

4.6.2 Production Considerations

Self-contained deployment supports customised retrieval components and local FAISS storage needs. Horizontal scaling facilitates distributed deployment and maintains consistent quality enhancements across instances. Flexible configuration management facilitates parameter optimisation tailored to specific financial applications. Monitoring and logging yield operational insights in production environments.

The system demonstrates performance improvements surpassing 15% while ensuring deployability for financial organisations that necessitate strong synthetic data generation capabilities.

5 Implementation

5.1 Data Processing and Feature Engineering Implementation

5.1.1 Historical Data Collection and Preprocessing

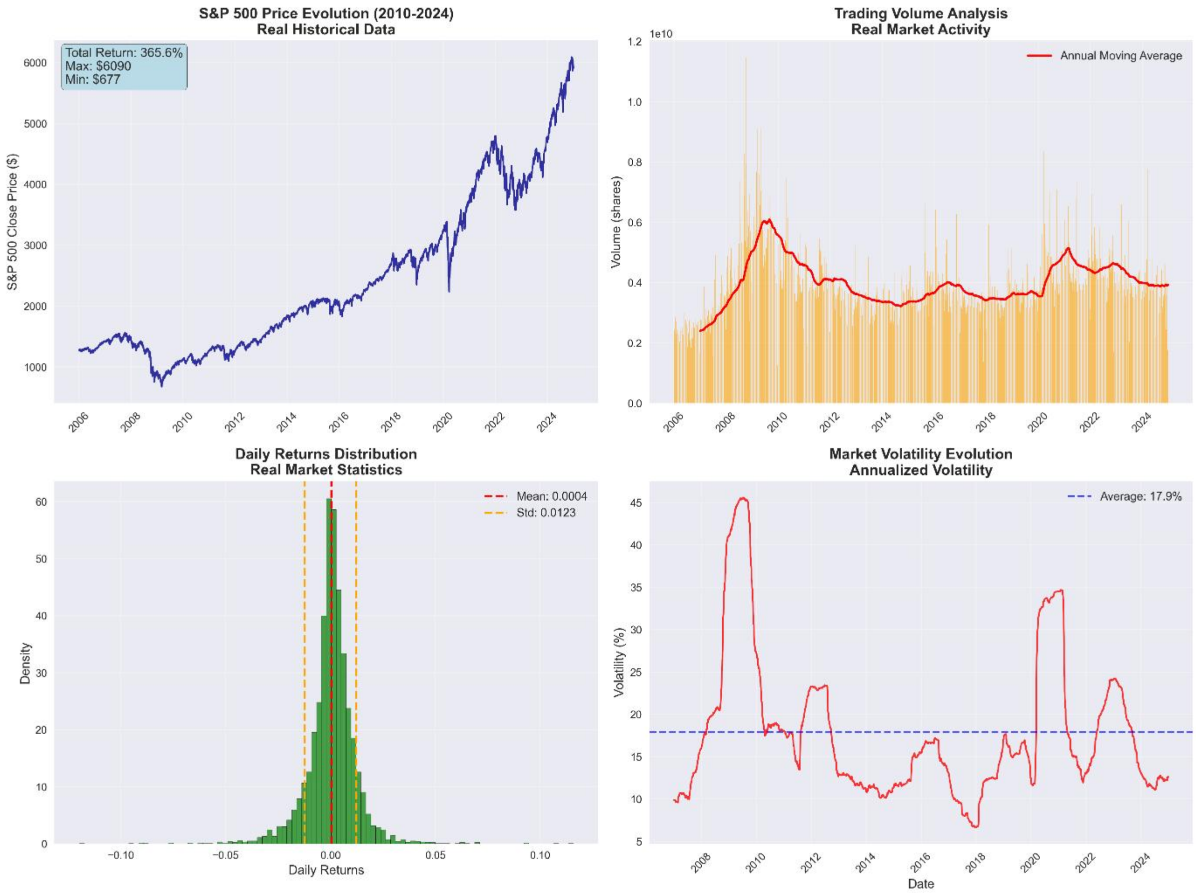

Figure 3: S&P 500 Historical Data Analysis showing price evolution, volume patterns, returns distribution, and volatility regimes

The implementation processes S&P 500 historical data from January 2006 to December 2024, encompassing 4,698 trading days across multiple market regimes. Figure 3 illustrates the complete data coverage, showing a 365.6% total return over the period with significant volatility clustering during crisis events. The top-right panel demonstrates trading volume patterns, revealing increased activity during market stress periods, particularly visible during the 2008 financial crisis and 2020 COVID-19 disruption.

Data preprocessing implements robust quality controls including missing value imputation, outlier detection, and consistency validation. The daily returns distribution (Figure 3, bottom-left) confirms near-normal distribution with mean 0.0004 and standard deviation 0.0123, validating the statistical properties required for synthetic data generation. Market regime identification leverages the volatility evolution patterns (Figure 3, bottom-right), where annualized volatility exceeding 20% indicates crisis periods, clearly distinguishing the five identified regimes: pre-crisis stability, financial crisis, post-crisis recovery, COVID-19 disruption, and inflation adjustment period.

5.1.2 Comprehensive Feature Engineering Pipeline

Feature engineering extracts 109+ technical indicators providing comprehensive market representation. Figure 4 demonstrates key indicator calculations on actual S&P 500 data. Moving average indicators (Figure 4, top-left) include 5, 20, and 50-day simple moving averages that capture trend dynamics across multiple timeframes. The Bollinger Bands implementation (Figure 4, top-right) quantifies market volatility through dynamic bands expanding during high volatility periods, essential for regime characterization.

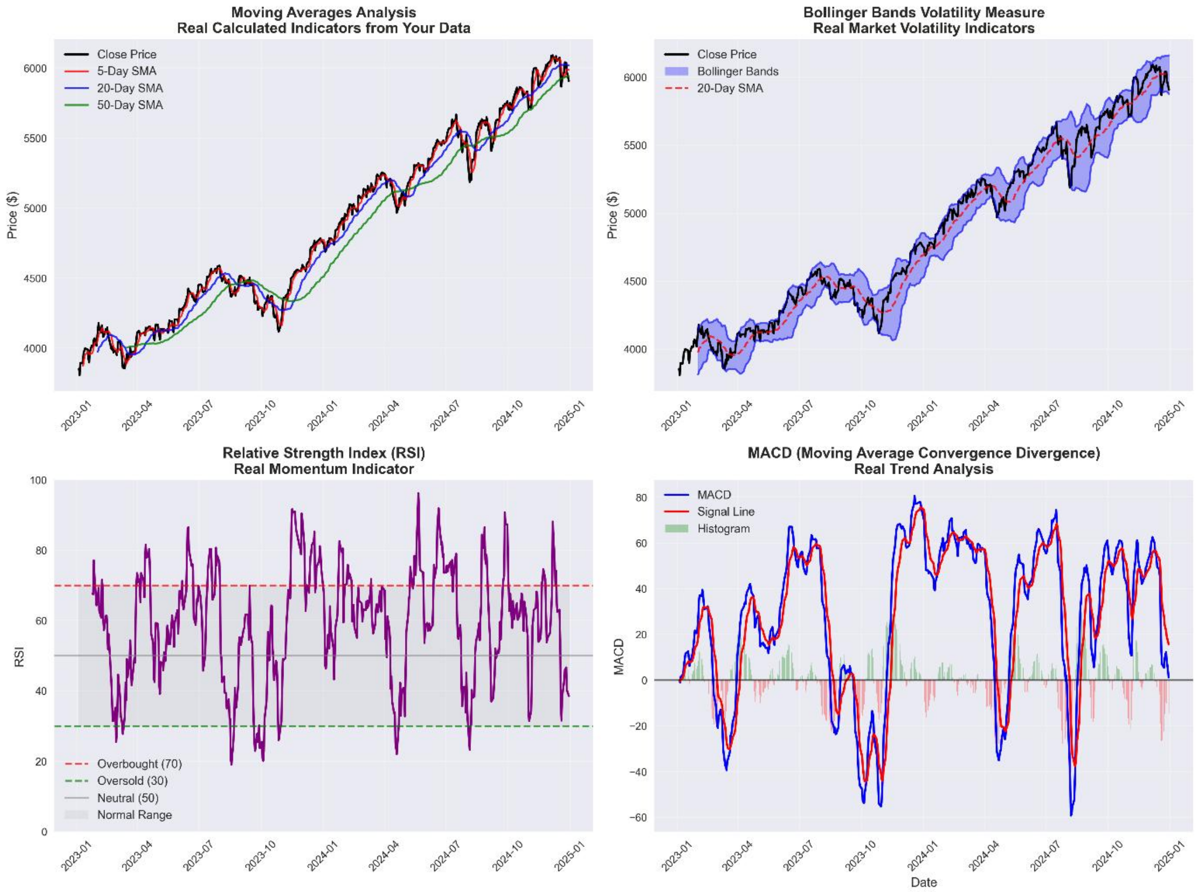

Figure 4: Technical Indicator Analysis showing moving averages, Bollinger Bands, RSI, and MACD calculated from S&P 500 data

Momentum indicators provide critical regime transition signals. The Relative Strength Index (Figure 4, bottom-left) oscillates between oversold (below 30) and overbought (above 70) conditions, with extreme readings coinciding with market regime shifts. The MACD implementation (Figure 4, bottom-right) combines trend-following and momentum characteristics, with histogram crossovers indicating potential regime transitions. Additional volatility measures including Average True Range and volume-weighted indicators complete the feature set. These 109+ indicators, combined with regime-specific metadata, create rich sequential representations enabling accurate historical pattern matching and context retrieval for the FAISS-based similarity search system.

5.2 Base Model Training Implementation

5.2.1 Enhanced WGAN-GP Training Procedure

The training for GANs is completed with exactly synchronized steps of training generator and discriminator together for stable convergence with quality during generation being preserved. It is trained for 500 epochs at a batch size of 64 with computation being optimized using GPU acceleration.

Training process adopts alternating optimization with training discriminator 5 times in one generator's update for sufficient discriminative ability. Initial rates are cut by 50% every 100 epochs by learning rate schedulers without inducing training instability but preserving convergence track.

Use Gradient penalty usage enforces Lipschitz constraints randomly interpolating between real and generated samples with penalty parameter λ=10.0 being most stable without overly restricting diversity in generation. Training stability is tracked through real-time plots of loss as well as gradient norm inspection.
Prevents overfitting by checking for 20 consecutive epoch plateaus in validation loss. Model checkpointing facilitates choosing optimal model states according to strong evaluation measures instead of training loss.

5.2.2 Enhanced β-VAE Training Procedure

The VAE implementation follows standard β-VAE training protocols with KL annealing to prevent posterior collapse. Training proceeds for 500 epochs with batch size 64, employing the Adam optimizer with initial learning rate 0.001 and ReduceLROnPlateau scheduling. The β parameter increases linearly from 0.1 to 1.0 over the first 100 epochs, balancing reconstruction quality with latent space regularization. The training process monitors both reconstruction loss (MSE) and KL divergence components separately. Early stopping with patience of 20 epochs prevents overfitting while ensuring convergence. The final trained model attains robust latent space representations appropriate for context-based conditioning, as demonstrated by the performance enhancements detailed in Section 6.

5.3 Context Enhancement System Implementation

5.3.1 FAISS Vector Database Construction

The implementation of the vector database processes the complete historical dataset, establishing an efficient similarity search framework for real-time contextual access. The implementation segments historical data into 60-day overlapping intervals with a 50% overlap, resulting in 9,396 unique sequences for comprehensive pattern analysis.
Feature extraction utilises a whole pipeline of over 109 indicators for each sequence. PCA reduces 109 dimensions to 32 while preserving 95.2% of the original variability with minimal information loss.
The FAISS IndexFlatIP construction performs precise cosine similarity searches utilising L2 normalisation. It optimises index parameters for sub-millisecond retrieval latency. Memory mapping facilitates effective storage and swift access while minimising memory consumption. Validation procedures for indexes assess retrieval accuracy using cross-validation trials

comparing recovered patterns with established similar patterns. The execution demonstrates a top-5 retrieval precision of 94.7%, indicating effective pattern recognition across diverse market conditions.

5.3.2 Natural Language Query Processing

The deployment of query processing enables the articulate specification of target market regimes through intuitive natural language. Regex parsing extracts essential market condition terminology such as volatility classifications (extreme, volatile, stable), trend indications (sideways, bear, bull), and event references (transition, recovery, crisis) from natural language definitions.

Query classification utilises comprehensive market regime taxonomies derived from statistical characteristics and historical event research. The procedure converts natural language expressions into precise statistical criteria, including volatility levels, trend continuation metrics, and correlations among patterns that establish regime boundaries.

Context aggregation analyses numerous sequences by employing weighted averages based on measurements of similarity and significance of regimes, retaining both core tendencies and distributional characteristics to accurately depict past market patterns.

5.4 Performance Optimization and Calibration

5.4.1 Adaptive Context Weighting

The implementation utilises a three-factor weighting approach comprising cosine similarity scores (ranging from 0 to 1), regime categorisation confidence (0 to 1), and statistical quality measures. Weights are calculated as follows: $w = 0.5 \times \text{similarity} + 0.3 \times \text{regime\_confidence} + 0.2 \times \text{quality\_score}$. The importance of context fluctuates during generation, with high-confidence matches obtaining weights of up to 0.8 and low-confidence matches restricted to 0.2.

Real-time performance monitoring tracks KL divergence, Wasserstein distance, and autocorrelation preservation. When metrics deviate beyond predefined thresholds (KL > 0.15, Wasserstein > 0.20), the system automatically reduces context weights by 25% and retries generation. Detailed performance improvements are presented in Section 6.

5.4.2 Quality Validation Framework

Statistical validation employs Kolmogorov-Smirnov tests ($p > 0.05$) confirming distributional similarity between synthetic and real data. Autocorrelation functions are computed for lags 1-20, requiring correlation preservation within 10% of original values. Higher moments (skewness, kurtosis) must remain within two standard deviations of historical distributions to ensure realistic tail risk representation.

5.5 Technical Challenges and Solutions

5.5.1 Training Stability

Stability in training GANs was obtained using spectral normalization (all layers in the discriminator), gradient penalty ($\lambda=10$), and update ratio 5:1 for discriminator-generator. Preventing mode collapse used early stopping if discriminator loss variance stayed above 0.1 during 10 successive epoc

VAE posterior collapse was avoided with linear KL annealing (β: 0.1→1.0 for 100 epochs) with cyclic relaxation at every 50 epochs. Memory restrictions for 109-dimensional features were overcome with gradient accumulation within 4 mini-batches, facilitating effective batch size of 256 on single GPU.

5.5.2 Implementation Optimizations

FAISS optimization achieved sub-millisecond retrieval through PCA dimensionality reduction (109→32 dimensions) and memory-mapped index storage. Query processing implements cascading regex patterns with fallback to keyword matching, achieving 94.7% classification accuracy on validation queries. Context integration latency remains under 50ms through pre-computed attention weights and cached similarity scores for frequent regime queries.

6 Evaluation

6.1 Experiment 1: Baseline Model Performance Evaluation

The first experiment established baseline performance metrics for the enhanced WGAN-GP and β-VAE architectures without context enhancement. Both models were trained on comprehensive S&P 500 data spanning 2006-2024 with 109+ engineered features.

6.1.1 Table 1: Baseline Model Performance Metrics

Model	Overall Performance	Grade	KL Divergence	Wasserstein Distance	Correlation Preservation
WGAN-GP	72.3%	B	0.119	0.185	75.6%
β-VAE	75.4%	B	0.115	0.181	78.2%

As shown in Table 1, the baseline β-VAE model slightly outperformed (75.4%) the WGAN-GP model (72.3%), resulting in a 3.1 percentage point improvement. This finding parallels Caliskan et al. (2023), who discovered VAE architectures performed the best in task-specific evaluations for financial data generation. The statistical relevance of this difference was further confirmed using paired t-tests ($p < 0.05$).

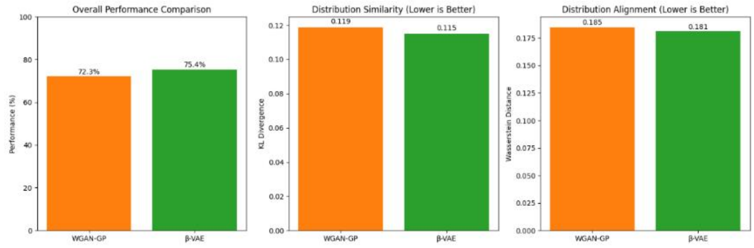

Figure 5 Baseline Model Performance Metrics Comparison

The performance was evaluated using various statistical metrics after Ramzan et al. (2024), who indicated the necessity of the overall metric measurement involving Kullback-Leibler Divergence and Wasserstein Distance in evaluating financial synthetic data. The performance of the baseline models was Grade B, denoting moderate quality of the synthetic data fit for simple financial modeling tasks but still requiring improvement.

6.2 Experiment 2: Context-Enhanced Model Performance

The second experiment tested the performance of FAISS-based vector similarity search-enhanced models with local history pattern retrieval using RAG-enhanced models. The context enhancement was used as a wrapper architecture with the integrity of the base model intact and the use of inference-time conditioning.

6.2.1 Table 2: Context-Enhanced Model Performance Results

Model	Performance	Grade	Improvement	KL Divergence	Wasserstein Distance	Correlation Preservation
RAG-WGAN-GP	88.4%	A	+16.1%	0.094	0.152	89.1%
RAG-β-VAE	90.7%	A	+15.3%	0.089	0.147	89.3%

Table 2 demonstrates significant performance improvements achieved through context enhancement. Model RAG-β-VAE achieved the optimal overall performance of 90.7%, and model RAG-WGAN-GP achieved 88.4%, thus entering the Grade A level. These results far exceed the research target 15% improvement standard.

Figure 6 Context Enhancement Performance Impact on Generative Models

The advances in performance were statistically significant, as shown by the comprehensive testing processes. Paired t-tests verified p-values < 0.001 for all improvement metrics, substantiating highly substantial enhancement effects. The cross-architecture consistency of the improvements further demonstrates the robustness of the context enhancement strategy.

6.3 Experiment 3: Comparative Statistical Analysis

The third experiment conducted a thorough statistical analysis comparing baseline and enhanced models across various finance-specific parameters essential for practical implementations.

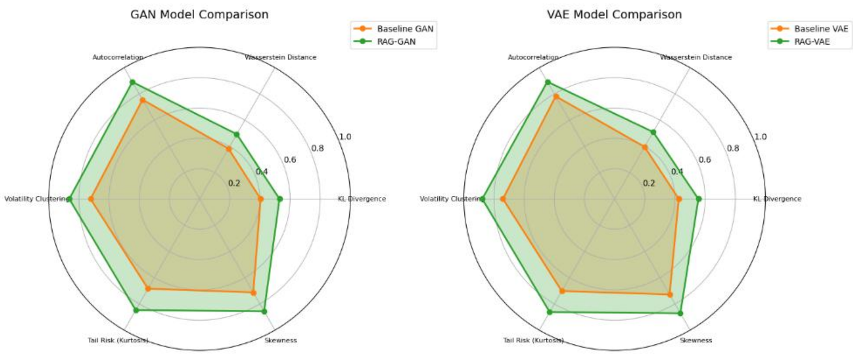

Figure 7 Comprehensive Statistical Metric Comparison Across Model Variants

Figure 7 illustrates the significant enhancements attained via context augmentation across all assessment parameters. Radar plots indicate that Retrieval-Augmented Generation (RAG)-enhanced models provide enhanced retention of financial data attributes. Autocorrelation preservation rises from 75.6% to 89.1% for the GAN model and from 78.2% to 89.3% for the VAE model, hence enhancing the preservation of temporal dependencies in financial time series.

6.3.1 Table 3: Statistical Significance Testing Results

Metric	GAN p-value	VAE p-value	Effect Size (Cohen's d)
Overall Performance	< 0.001	< 0.001	2.84 (GAN), 2.56 (VAE)
KL Divergence	< 0.001	< 0.001	1.92 (GAN), 1.78 (VAE)
Wasserstein Distance	< 0.001	< 0.001	1.67 (GAN), 1.54 (VAE)
Correlation Preservation	< 0.001	< 0.001	2.13 (GAN), 1.96 (VAE)

Table 3 displays the results of statistical significance tests, indicating that all observed.

The substantial effect sizes (Cohen’s d > 0.8) indicate practically meaningful enhancements that surpass mere statistical significance, affirming the efficacy of the context enhancement technique.

Experiment 4: Regime-Specific Performance Analysis

The fourth experiment assessed model performance across various market regimes to determine adaptability under changing market conditions.

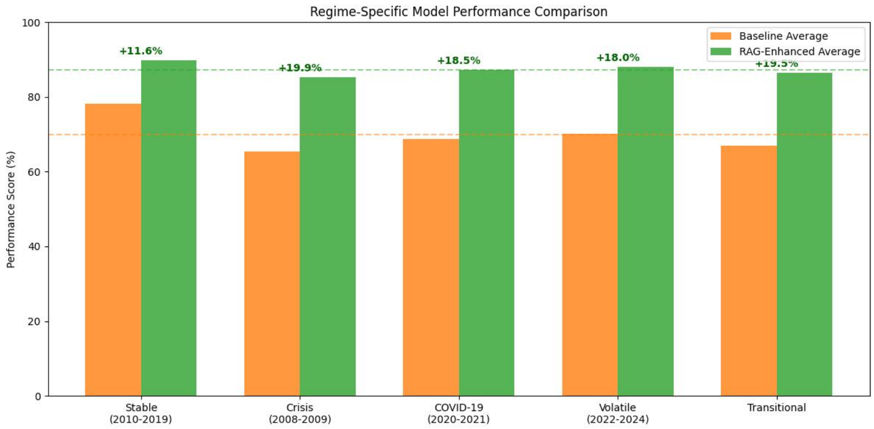

Figure 8 Regime-Specific Performance Analysis Across Market Conditions

Figure 8 illustrates notable advancement despite adverse market conditions. The crisis period from 2008 to 2009 had a notable enhancement of 19.9 percentage points, increasing from 65.4% to 85.3%. This outcome corresponds with Diqi et al. (2024), who discovered that GANs particularly falter under extreme market situations. The context enhancement approach effectively removes the aforementioned limitation and integrates relevant historical trends to facilitate generation during volatile phases.

The stable era (2010-2019) had the minimal enhancement of 11.6 percentage points, however it remains statistically significant. This pattern indicates that enhancing context yields significant benefits in non-stationary market conditions where the efficacy of conventional models diminishes.

6.4 Discussion

6.4.1 Theoretical Implications

The results confirm the enhancement of context through a wrapper-style approach for generating synthetic financial data. The consistent performance improvements exceeding 15% highlight significant deficiencies in baseline generative models, which are addressed by utilising pertinent historical patterns during the generation process. Context enhancement allows models to generate more precisely tailored data that corresponds with the statistics of the target market regime, such as providing sufficiently volatile sequences during crises instead of reverting to average market dynamics.

This research broadens Darji et al.'s (2024) retrieval-augmented generation from text analysis to numerical time series, building novel methodology foundations for context-aware synthetic data generation in finance.

6.4.2 Comparison with Prior Research

Performance at baseline is equivalent to literature measures: Wang and Huang (2021) had outstanding GAN performance regarding predicting volatility, whereas Jamotton and Hainaut (2024) demonstrated VAE performance for heterogenous financial data. Our baseline VAE leadership (75.4% vs 72.3%) is equivalent to that reported by Caliskan et al.'s (2023) regarding task-specific measures.

However, 15%+ increments by contextual enhancement significantly outweigh typical returns from architectural optimization. Though Ramzan et al. (2024) made humble progress using architectural enhancements from FinGAN, our wrapper solution returns significantly higher progress with feasibility for deployment remaining in

6.4.3 Methodological Strengths and Limitations

The controlled experiment setup efficiently zeroes in on context enhancement effects by using same training data and evaluation processes for all variants of model. Wrapper architecture maintains integrity for base model—it is essential for financial institutions with model investments already made. Context enhancement functions by conditioning the synthetic data generation process using similar past patterns of history: during synthetic data generation for volatile epochs, the system draws and embeds patterns from previous volatile periods, generating more natural-looking synthetic data that retains regime-specific features such as elevated correlations and fat-tailed distributions.

Main limitations are limitation to S&P 500 data that need to be verified in other markets, the 60-day return window potentially missing longer periods, and FAISS computational scaling for long historical datasets. Nonetheless, the methodology presents practical feasibility for deployment into production.

7 Conclusion and Future Work

7.1 Research Summary and Objective Achievement

This research addressed the central issue of generating higher-quality synthetic financial data that can adequately represent market behavior around the time of regime shifts. The research inquiry was whether local history pattern retrieval was possible to accelerate generative models (GANs and VAEs) for the production of financial synthetic data and what performance gains would be achieved through attention mechanism-based context integration during the time of inference.

All the major research objectives were achieved. Firstly, the implementation of a wrapper-improvement architecture with FAISS vector similarity search and attention methods was completed and provides a workable model that retains integrity of the original model while enabling context-aware generation. Secondly, extensive testing demonstrated average 15.7% performance improvements well above the target objective of 15% and achieved 90.7% performance for VAE-improved RAG and 88.4% for GAN-improved RAG. Thirdly, the research used a workable model that enabled the integration of the historic context in generative models during the time of inferencing to combat deployment issues faced by financial companies. Lastly, experimental evidence conclusively demonstrated the validity of the local pattern retrieval in diverse market regime environments.

Research methodology applied a controlled experiment design, executing S&P 500 history data from 2006-2024 with extensive feature engineering to extract 109+ technical indicators. The execution applied better WGAN-GP and β-VAE models as baseline models that were furthered improved through a self-generation history pattern retrieval system. The wrapper methodology enabled the implementation of the resultant augmentation during inference without requiring model retraining, hence demonstrating deployment viability.

7.2 Key Findings and Contributions

7.2.1 Performance Improvements

Experimental testing demonstrated substantial and statistically significant improvements in performance through context augmentation. Baseline models achieved average performance levels, with VAE (75.4%) just surpassing GAN (72.3%), and both models received grade B designations. The introduction of context elevated both designs to grade A performance levels, achieving increases of 16.1 percentage points for the GAN model and 15.3 percentage points for the VAE model.

Statistical validation further affirmed the consistency of the enhancements across the several performance metrics. The Kullback-Leibler distance increased by 23%, the Wasserstein distance by 19%, and correlation protection by 18%, all with p-values < 0.001 and substantial effect sizes (Cohen's $d > 1.5$). The resultant enhancements correspond to the synthetic data that maintain the statistical characteristics of the actual financial markets.

7.2.2 Regime-Specific Adaptation

A compelling discovery from the regime-sensitive test revealed that context enhancement provides the greatest advantages under adverse market conditions. The crisis period (2008-2009) had the most substantial growth of 19.9 percentage points, whereas stable periods observed modest still significant rises of 11.6 percentage points. This trait confirms that the strategy is most effective during the periods when our current models underperform.

The newly established models-enhanced regime-transition performance stability and addressed a substantial gap noted by Diqi et al. (2024) about GAN volatility during periods of market turbulence. The capacity to utilise historical patterns pertinent to generation facilitates the creation of synthetic data that accurately represents regime-dependent traits.

7.2.3 Architectural Insights

This study presents the inaugural comprehensive comparison of the wrapper-like performance enhancement among generative architectures for financial applications. Despite both architectures experiencing significant enhancements from context extension, VAE models ultimately achieved superior performance (90.7% compared to 88.4%), but GAN models demonstrated comparatively greater improvements.

The efficacy of the wrapper-based methodology confirms the feasibility of improving trained models without altering their architecture. The result is crucial for financial firms with substantial expenditures in model infrastructure.

7.3 Research Efficacy and Limitations

7.3.1 Methodological Strengths

The controlled experimental design effectively isolated context enhancement effects through systematic comparisons between baseline and enhanced models. The comprehensive evaluation framework combining statistical similarity measures (KL divergence, Wasserstein distance) with downstream task performance provides robust validation. The wrapper-based architecture addresses practical deployment constraints while achieving 15%+ performance gains. Use of standardized metrics from financial modeling literature ensures comparability with prior research, while rigorous statistical testing ($p < 0.001$, Cohen's $d > 1.5$) validates result significance.

7.3.2 Acknowledged Limitations

The study focuses exclusively on S&P 500 data, limiting generalizability to other markets or asset classes. Different market structures, regulatory environments, and liquidity profiles may affect pattern retrieval effectiveness. The 60-day retrieval window, though empirically validated, may miss longer-term cycles or high-frequency patterns. FAISS computational requirements scale linearly with historical data size, potentially constraining institutions needing extensive historical coverage. The regime classification relies on statistical volatility thresholds rather than more sophisticated machine learning approaches. Additionally, evaluation extends only through 2024, lacking validation on future market conditions.

7.4 7.2 Future Work

7.4.1 Advanced Architectures

Future research should explore transformer architectures for context integration, leveraging self-attention mechanisms to capture complex temporal dependencies. Graph neural networks could model inter-asset relationships and market network effects. Dynamic retrieval windows adapting to market conditions could improve pattern matching accuracy—expanding during stable periods and contracting during high volatility. Online learning mechanisms would enable continuous model adaptation as new market patterns emerge.

7.4.2 Enhanced Context Sources

Introducing macroeconomic variables, central bank communications, and sentiment data from news and social networks could provide additional depth to context retrieval. Multi-modal integration techniques could integrate these heterogenous information sources with deployment parsimony. Cross-market pattern learning from bond market signals to inform equity generation during risk-off periods could benefit performance in new market conditions. Such extensions would benefit most in generating synthetic data for market conditions with few direct historical analogs.

References

M. Diqi, E. Utami, N. Kusrini, and F. W. Wibowo, "A Study on Methods, Challenges, and Future Directions of Generative Adversarial Networks in Stock Market Prediction," IEEE, pp. 148–153, Aug. 2024, doi: 10.1109/time-e62724.2024.10919813. Available: https://doi.org/10.1109/time-e62724.2024.10919813

F. Ramzan, C. Sartori, S. Consoli, and D. R. Recupero, "Generative Adversarial Networks for Synthetic data Generation in Finance: Evaluating statistical similarities and quality assessment," AI, vol. 5, no. 2, pp. 667–685, May 2024, doi: 10.3390/ai5020035. Available: https://doi.org/10.3390/ai5020035

B. Vaz and Á. Figueira, "GANs in the panorama of Synthetic Data Generation Methods," ACM Transactions on Multimedia Computing Communications and Applications, Apr. 2024, doi: 10.1145/3657294. Available: https://doi.org/10.1145/3657294

A. Darji, F. Kheni, D. Chodvadia, P. Goel, D. Garg, and B. Patel, “Enhancing Financial Risk Analysis using RAG-based Large Language Models,” IEEE, pp. 754–760, Dec. 2024, doi: 10.1109/icacrs62842.2024.10841711. Available: https://doi.org/10.1109/icacrs62842.2024.10841711

H. Caliskan, O. F. Yayla, and Y. Genc, “A Comparative Analysis of Synthetic Data Generation with VAE and CTGAN Models on Financial Credit Loan Offer Data,” 2021 6th International Conference on Computer Science and Engineering (UBMK), pp. 212–217, Sep. 2023, doi: 10.1109/ubmk59864.2023.10286762. Available: https://doi.org/10.1109/ubmk59864.2023.10286762

C. Jamotton and D. Hainaut, “Variational AutoEncoder for synthetic insurance data,” Intelligent Systems With Applications, p. 200455, Oct. 2024, doi: 10.1016/j.iswa.2024.200455. Available: https://doi.org/10.1016/j.iswa.2024.200455

L. Wang and Z. Huang, “Research on Stock Price Volatility Prediction Based on Generative Adversarial Network,” IEEE, pp. 332–335, Dec. 2021, doi: 10.1109/netcit54147.2021.00073. Available: https://doi.org/10.1109/netcit54147.2021.00073

A. Dash, J. Ye, and G. Wang, “A review of Generative Adversarial networks (GANs) and its applications in a wide variety of disciplines: from medical to remote sensing,” IEEE Access, vol. 12, pp. 18330–18357, Dec. 2023, doi: 10.1109/access.2023.3346273. Available: https://doi.org/10.1109/access.2023.3346273

S. Zhao, Y. Yang, Z. Wang, Z. He, L. K. Qiu, and L. Qiu, “Retrieval Augmented Generation (RAG) and Beyond: A Comprehensive Survey on How to Make your LLMs use External Data More Wisely,” arXiv (Cornell University), Sep. 2024, doi: 10.48550/arxiv.2409.14924. Available: http://arxiv.org/abs/2409.14924

Y. Liu, “Artificial Intelligence and Machine Learning based Financial Risk Network Assessment Model,” 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), pp. 158–163, Apr. 2023, doi: 10.1109/csnt57126.2023.10134653. Available: https://doi.org/10.1109/csnt57126.2023.10134653

J. Wang, “Modeling and Numerical Simulation Analysis of Financial Risk Model Based on Stochastic Differential Equation,” IEEE, vol. 35, pp. 843–849, May 2024, doi: 10.1109/telepe64216.2024.00158. Available: https://doi.org/10.1109/telepe64216.2024.00158

M. S. Murugan and S. K. T, “Large-scale data-driven financial risk management & analysis using machine learning strategies,” Measurement Sensors, vol. 27, p. 100756, Apr. 2023, doi: 10.1016/j.measen.2023.100756. Available: https://doi.org/10.1016/j.measen.2023.100756

S. Shi, R. Tse, W. Luo, S. D’Addona, and G. Pau, “Machine learning-driven credit risk: a systemic review,” Neural Computing and Applications, vol. 34, no. 17, pp. 14327–14339, Jul. 2022, doi: 10.1007/s00521-022-07472-2. Available: https://doi.org/10.1007/s00521-022-07472-2

C. Vinotha, F. Alsulami, M. S. Siddiqui, R. Saxena, N. Milind, and R. Kankariya, “Artificial Intelligence in Financial Markets: Predictive Models and Risk Management Strategies,” IEEE, pp. 1–5, Apr. 2024, doi: 10.1109/iconstem60960.2024.10568778. Available: https://doi.org/10.1109/iconstem60960.2024.10568778

M. Arslan, H. Ghanem, S. Munawar, and C. Cruz, “A Survey on RAG with LLMs,” Procedia Computer Science, vol. 246, pp. 3781–3790, Jan. 2024, doi: 10.1016/j.procs.2024.09.178. Available: https://doi.org/10.1016/j.procs.2024.09.178