

Proposal: Quantitative Trading Framework for Cryptocurrency Market in Futures Modality

Table of Contents

1. [Introduction](#)
 2. [Background](#)
 3. [Proposal](#)
 1. [Data Ingestion](#)
 2. [Data Observation and Transformation](#)
 3. [Modeling](#)
 4. [Results Analysis](#)
 5. [Framework Output](#)
 4. [Conclusions](#)
-

Introduction

This document presents a comprehensive proposal for developing a quantitative trading framework specifically tailored for the cryptocurrency market, specifically in the futures modality. The proposal is made from the point of view of Data Science, leveraging machine learning and time series analysis, based on a robust, data-driven approach.

First, the document explains the problem and the context of this complex scenario. Then it moves to define the proposal, including bibliographic support for specific recommendations and case-based analysis for details that depend on the results of the implementation.

Background

Quantitative trading is a method that uses mathematical, computational and statistical models to identify trading opportunities. It is a data-driven strategy that focuses on quantitative analysis from vast datasets to generate trading insights and decisions ^{^1}.

The **cryptocurrency market** consists of digital or virtual currencies that use cryptography for security and operate on decentralized networks, typically based on blockchain technology. This market is defined by several key characteristics that differentiate it from traditional financial markets ^{^2}.

It operates 24 hours a day, 7 days a week, without any market closures. It is also known for its high volatility, with significant price fluctuations occurring over very short time horizons, presenting both unique opportunities and risks for traders. Furthermore, the market provides unprecedented transparency through publicly available blockchain data, allowing for the analysis of on-chain metrics ^{^2 ^3}.

In the volatile crypto market, algorithms can process vast amounts of historical data and perform complex analyses to provide projections, offering a significant advantage where timing is critical.

A **futures contract** is a standardized legal agreement to buy or sell an underlying asset at a predetermined price and specified future date. In the cryptocurrency context, futures trading allows for speculation on price movements without directly owning the asset. A dominant instrument is the perpetual swap, a futures contract with no expiry date ⁴. Its price is pegged to the underlying spot price through a periodic funding rate mechanism, where longs and shorts exchange payments based on the price difference.

Futures trading employs sophisticated margin mechanisms with important risk considerations. Initial margin requirements typically range from 2% to 50% (50x to 2x leverage), with maintenance margin levels triggering liquidation warnings. The choice between cross-margin and isolated margin modes represents a critical risk management decision.

Isolated margin allocates margin to individual positions independently, limiting loss to allocated margin and preventing cross-position contamination, but with lower capital efficiency and potential for underutilization. This mode suits new strategies, high-risk positions, and experimental approaches. Cross margin uses a shared margin pool across all positions, offering higher capital efficiency and better margin utilization, but carrying contagion risk and potential for cascading liquidations. This mode works better for established strategies and portfolio-level risk management.

Proposal

1. Data Ingestion

The data collection strategy must differentiate between model development/training phases and live production usage.

For model development, batch data collection requires comprehensive representation of all operational scenarios with balanced representation across different cryptocurrencies, trading pairs, and market conditions. Historical data should span multiple years (minimum 3-5 years) to capture seasonal patterns, market cycles, and long-term trends. Multi-frequency data must be properly labeled to enable training across different time horizons, and data should include diverse market regimes with appropriate representation.

For real-time production usage, continuous data streams in the same format as training data are essential. This data serves both predictions and ongoing model performance monitoring, with automated alerting systems for special market events and model performance degradation integrated with monitoring systems for operational oversight.

Implementation details

For the data collection, the pipeline should perform a regular data extraction from major cryptocurrency exchanges like Binance, OKX, and Bybit with configurable extraction frequency and adaptive scheduling during high volatility periods. The pipeline should support multiple data types including OHLCV, order book data, funding rates, and open interest.

The data should contain meta-parameters for enhanced analysis beyond the prices fluctuation. Including exchange identifiers, liquidity metrics, asset classifications, market capitalization tiers, regulatory jurisdiction, etc.

or real-time infrastructure, robust WebSocket clients with exponential backoff retry logic, message queuing systems for handling data bursts, parallel processing for high-frequency data streams, and comprehensive monitoring for data completeness and quality are essential, with automated alerting integrated into model validation systems.

2. Data Observation and Transformation

This section focuses on data processing on the training and model selection phase. The transformations determined at the end of the first iteration of training will be integrated to the real-time usage of the framework.

Since the dataset is large and varied, it is necessary to use automated tools of data observation. Is necessary to include automatic continuity analysis to detect data gaps, statistical outlier detection using Z-score and domain-specific thresholds, validation of logical consistency, and timestamp monotonicity; then analyze those cases manually.

The data needs to be cleaned, using multiple approaches depending on the problems detected. For example, data removal for irrecoverable corruption, forward-filling for short interruptions, linear interpolation for medium gaps, probabilistic imputation for complex multivariate missing data, and multi-source validation using alternative data providers for critical gaps.

Then the data needs to be normalized. In this type of data the main concerns are timestamp synchronization across all sources, currency normalization to stable references, volatility-adjusted scaling for price-based features, and robust scaling techniques to handle fat-tailed distributions.

If we have categorical data it must be encoded using the right method according the type of feature; either Label Encoding, Ordinal Encoding, One-Hot Encoding, or others.

Feature analysis involves correlation analysis to identify redundant features, mutual information scoring for non-linear relationships, recursive feature elimination with cross-validation, and principal component analysis for dimensionality reduction in highly correlated feature sets.

Data split

After the data is cleaned and properly transformed is necessary to determine how to split the data to perform training on several models with variety of hyper-parameters, then test on the chosen metrics, and adjust the parameters, while leaving out a representative group of data, never seen by any model, to perform the final model validation and selection.

Standard cross-validation methods, particularly K-fold approaches, are not appropriate for financial time series. It is necessary to keep in mind the inherent temporal dependencies, to avoid future information to influence past predictions.

The proposed strategy is to perform a walk-forward strategy. Defining a large training window encompassing the initial 80-90% of chronological data. Then a validation window containing the last portion of the time-series that must be kept separated from the training set until the end of the modeling process. The large training window then can be split in sliding windows of train-test segments during the models training.

3. Modeling

Baseline and Candidate Models

It is recommendable to have a baseline "model" as reference, for this framework an ARIMA model ^{^5} can be used as comparison.

Primary model candidates based on literature review include Gradient Boosting Machines like LightGBM and XGBoost, LSTM networks variants for temporal dependency capture, CNN-LSTM hybrids for multi-scale pattern recognition, and more recently Transformer networks for long-range dependency modeling ^{^6 ^7}.

Recent research indicates that Conv-LSTM architectures, bi-directional LSTM, and properly regularized XGBoost Regressor typically demonstrate strong performance in cryptocurrency forecasting tasks.

Evaluation Metrics

The more frequent metrics in the literature, for the proposed model, at training are Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2).

But in the validation stage more context specific metrics should be included. Like Sharpe Ratio and Sortino Ratio, drawdown analysis including maximum drawdown and recovery period, profitability measures such as profit factor and win rate, and strategy efficiency metrics like gain-to-pain ratio.

Cryptocurrency-specific metrics include funding rate impact on strategy performance, liquidation risk exposure and margin efficiency, regime-aware performance across different market conditions, and transaction cost viability after realistic cost accounting.

Cross-validation training and hyper-parameters adjustment

It is proposed to used a sliding window approach, based on a fixed-size training window that moves forward with each step; this reduces the data used in training and discards potentially valuable historical data, but forces the model to adapt to recent market conditions.

The size of the sliding window can be a hyper-parameter to adjust, ranging from 1 month to 3 or 6 months, depending on the dataset and models restrictions.

If for some models the combination of possible/relevant hyper-parameters is too large to perform a training round for each combination, then based on a literature revision a subset must be determined, using recommended values that have scored high in the metrics that are of interest in this framework, and using a similar dataset.

Model Selection for Production

The multi-criteria evaluation framework primary ranks by out-of-sample financial metrics like Sharpe ratio and maximum drawdown, with secondary assessment of statistical performance, tertiary evaluation of operational feasibility, and quaternary analysis of interpretability and failure mode understanding.

Robustness testing involves performance consistency across multiple validation periods, sensitivity analysis to hyperparameter variations, stress testing under extreme market conditions, and capacity analysis with increasing position sizes.

Ensemble and Combination Strategies

A strategy that is rarely mentioned in the literature is to create a dynamic ensemble architecture. This allows to use models that excel in different circumstances Using several models that perform different on several factors, so we can train a layer that learns to determine which output (which model) is the best for each input.

Ensemble method options include static weighted average based on historical performance, dynamic weighting using simple meta-models, model stacking with out-of-fold predictions as features, and Bayesian model averaging for probabilistic combination.

4. Results Analysis

This section presents the analysis outcomes from the modeling work.

Model Performance Comparison: The validation results from walk-forward testing will show each model's performance across different market conditions. This analysis will determine which models demonstrate consistent risk-adjusted returns and identify their specific strengths in various market regimes.

The analysis will reveal which features drive predictive power across different models. Technical indicators like volatility measures and momentum oscillators may show varying importance depending on market conditions, while on-chain metrics like exchange flows could demonstrate predictive value during regime transitions.

The ensemble performance analysis will determine whether combining models provides benefits over individual approaches. This includes assessing whether certain models complement each other by performing well in different market environments or if a single model consistently outperforms others.

Based on these results, decisions will be made about model selection for production deployment. If certain models show unexpected failure modes or if new patterns emerge in the validation, it may be necessary to perform additional training rounds with adjusted feature sets or model architectures.

The key outcome of this analysis is determining the optimal configuration for live trading, whether that involves a single model, a static ensemble, or a dynamic ensemble that adjusts based on market regime detection.

5. Framework Output

At the final stage is time to determine the selected model or ensemble configuration ready for deployment, including all necessary parameters and preprocessing components.

The framework must also determine te specific rules for generating trading signals based on model predictions, including confidence thresholds and position sizing parameters derived from validation performance.

As we mentioned earlier, is recommended to develop a monitoring framework. Including performance benchmarks and alert thresholds for detecting model degradation, plus automated retraining triggers based on the validation methodology.

Finally the framework must include the documentation, complete specifications for integrating it into a live trading environment, including data requirements and execution protocols.

Conclusions

This proposal has outlined a comprehensive framework for developing a quantitative trading system specifically designed for cryptocurrency futures markets.

The framework establishes a complete pipeline from data collection through model deployment, with particular emphasis on robust validation techniques suitable for financial time series. The walk-forward validation methodology and multi-dimensional evaluation metrics provide a solid foundation for assessing model performance in realistic market conditions.

The proposed modeling approach offers flexibility through multiple candidate architectures and ensemble strategies, allowing the system to adapt to different market regimes and capitalize on diverse predictive patterns.

Future enhancements could explore additional data sources, advanced ensemble methods, and more sophisticated regime detection algorithms. The modular design of the proposed framework allows for ongoing development and refinement as new techniques and data become available.