School of Economics, Business and Computer Science

AI Based Real Time detection of Blockchain Market Anomalies

Danila Periakov
Dr. Marios Touloupou
2026

School of Economics, Business and Computer Science

AI Based Real Time detection of Blockchain Market Anomalies

Diploma Thesis submitted for Bachelor of Science in Applied Computer Science at Neapolis University Paphos

Danila Periakov
Dr. Marios Touloupou
2026

THE SOLEMN DECLARATION

I, **Danila Periakov**, am aware of the consequences of plagiarism, I declare responsibly that this work, entitled **" AI Based Real Time detection of Blockchain Market Anomalies"**, is the product of strictly personal work and all the sources I have used have been properly stated in the bibliographic references. The points where I have used ideas, text and/or sources of other authors, are clearly mentioned in the text with the appropriate citation and the relevant reference is included in the section of the bibliographic reports with a full description.

Danila Periakov

Table of Contents

Abstract

This research proposes an artificial intelligence (AI)-based real-time anomaly detection system that perceives anomalies in Bitcoin perpetual futures markets. The limitation of traditional static monitoring tools resolved in the proposed framework as static thresholds cannot differentiate between a true anomaly or a candle that does not strictly fall within a given range. In trending stable market conditions, a threshold-based technique fires on normal candles resulting in alert fatigue instead of actionable signals.The proposed monitoring system makes use of the Isolation Forest algorithm, which is an unsupervised machine learning algorithm that detects anomalous behaviour in price and volume time series without any labelled training data.

The architecture consists of three coupled but functionally independent analytical modules: Price/Volume Module featuring Isolation Forest and Order Flow Imbalance analysis, Liquidation Heatmap Proxy based on open interest public data, and Funding Rate Module based on perpetual future funding mechanism. Each module generates an independent directional probability estimate on a scale. The final UP/DOWN/NEUTRAL signal model is a weighted average (with weights easily verifiable) of sub-module signals. The model is based solely on data obtained from the Binance Futures public REST API documenting the reproducibility of the model.

The prototype was demonstrated with a live simulation for a period of two weeks using the BTCUSDT perpetual futures pair. It was noted that the Isolation Forest model demonstrated the ability to adapt to changing volatility regimes, the Order Flow Imbalance indicator was capable of providing responsive trend indications, and significant events such as a spike in volume, extreme funding rates and rise in volatility were also successfully detected.

During live testing over the course of two weeks, the system has successfully discerned three distinct types of events: identified volume spike with no price movement, identified extreme funding rate in conjunction with price heatmap bearish imbalance, and identified “Anomalous” volatility regime with OFI at +0.11 — these events would not have been able to be successfully identified in a fixed-threshold machine learning system without the confounding occurrence of many false-positive events.The full source code for the system proposed in this thesis is publicly available in GitHub. The repository includes all detection modules, the developed Streamlit dashboard and requirements.txt fulfilling the requirements to create the same running environment. The system has no secret API keys, paid services or additional computing power required.

Repository: https://github.com/Numpack/Thesis

Introduction

Background of Financial and Digital Markets

The landscape of financial markets has changed dramatically over the past two decades. Advances in technology have led to algorithmically processed high-frequency trading through exchanges with geographic separation in milliseconds. Equity and commodity markets are now increasingly facing new competition from digital asset markets that trade perpetually without typical business hours or the circuit-breakers used to mitigate volatility in traditional financial markets.

Perpetual futures contract is one of the most actively traded instruments in cryptocurrency derivative trading markets. It allows participants to gain leverage into bitcoin exposure without expiry date. Offered by exchanges such as Binance, Bybit, and OKX, participants are 'longed' or 'shorted' into bitcoin price which causes the unique effect of funding rate payments, leading to increasingly growing open-interest, and the risk of cascading liquidations upon mass closings of long or short positions. By 2024, perpetual futures trading of bitcoin on Binance accounted for more than half of global volumes, and was the most popular venue for bitcoin (CoinMarketCap, 2024).

The continuous trading hours, highly leveraged positions, and mainly retail investor base make cryptocurrency futures an attractive candidate for potential market anomalies. Volume bursts, abrupt regime shifts, and extreme funding rates are events that can happen instantaneously, and quickly transmit across positions. For this reason, the ability of market participants and academics studying digital asset market microstructure to identify these occurrences in real time can be rather valuable.

Problem Statement

The current methods of market monitoring typically involve the application of static statistical methods with simple threshold-based alerts. The methodologies have certain inherent weaknesses that restrict their ability to detect anomalous events of sudden onset in real time.

The current systems include:
1. rule-based thresholds,
2. lagged reporting of sudden changes in volatility,
3. limited consideration of data from other related markets and
4. lack of learning algorithms to separate typical market variability from truly anomalous events.

These limitations would be overcome by AI based techniques that are adapted to the market environment and not fixed by rules as mentioned in this study.

Research Aim and Objectives

The primary aim of this research is to explore the use of Artificial Intelligence in real-time monitoring and anomaly detection for improving situational awareness in financial markets. It proposes to investigate the real-time identification of abnormal market behavior that could lead to the early generation of alerts for risk-sensitive decision-making. The present research does not intend to investigate price prediction. If the system generates a directional probability — that is, UP or DOWN likelihood — it implies that the market is in the asymmetrical state at the moment — either there are more liquidation zones lying above the current price than below or vice versa, or net buyers are more aggressive than net sellers or vice versa, or there is more leveraged buying than selling or vice versa. This is simply a characterisation of the current market asymmetry and there is no intention to predict which way the price will go in the future.

The research objectives are:

To explore Artificial Intelligence techniques for continuous real-time monitoring of price, volume, liquidity, and related market indicators.

To examine the state of the art techniques of anomaly detection in financial markets for real-time alert generation.

To analyze the shortcomings of the existing real-time monitoring systems that utilize static threshold or simple algorithms.

To examine the patterns and anomalies that represent abnormal behavior in financial markets without forecasting prices.

To evaluate how an Artificial Intelligence based conceptual model can contribute towards enhancing situational awareness for risk managers.

Research Questions

This proposal studies the theoretical underpinnings, evidence, and information value of AI techniques used to identify anomalies in financial markets. The three research questions outlined below define the scope of the study, which is, respectively, about methodology, data properties, and the value of real-time alerts to inform market understanding.

RQ1.
What are the theoretical and methodological differences between AI-based techniques used to identify anomalies in financial markets and traditional techniques used for market surveillance, how do these techniques respond to new or changing market conditions in various market settings?

This question examines the theoretical and methodological differences and responsiveness of the techniques under investigation.

RQ2.

What properties of market data such as intraday trading volume, order-book depth and shape, liquidity, order size distribution, price distribution, and order flow imbalance are most informative for identifying anomalous market events when analyzed on a continuous and real-time basis in various market settings?

This question addresses the relative informativeness of the data properties identified above for anomaly detection when analyzed on a continuous and real-time basis in the techniques under investigation.

RQ3.
How does real-time notification of detected market anomalies contribute to understanding of current market conditions and how may this improved understanding of market conditions influence the perception of market dynamics during periods of high volatility in various market settings?

This question considers how real-time notification of detected anomalies may improve the understanding of market conditions without specifying a response to improved understanding in the techniques under investigation.

Scope and Limitations

The only market being simulated in this study is the perpetual futures contract for BTCUSDT. The only exchange implemented into the prototype is Binance. The scope has been constrained in this manner in order to achieve a prototype that is technically sound, while remaining explanatory in scope, and allowing other researchers to reproduce the findings. The system does not make provision for trading on any other cryptocurrency or cryptocurrency counts, nor does it implement the spot market or other markets for securities and traditional assets.

This thesis exclusively relies on the Binance Futures public REST API. Access to candlestick data, open interest, and funding rate is available, while order book depth, trades execution, and any position information which the exchange may keep are not accessible. Therefore, the Liquidation Heatmap is only served as an estimate.

This was a prototype system build with research purpose so it is not intended to be used in production or actual market for actual trading/execution or for making production decision. The directional probability as an output is interpreted as the current asymmetry in the market, it is not a price prediction. Observation method is used to evaluate the system during a 2-week active period . There is no known labelled ground truth dataset for the active anomalous events in this market. Parameter tuning was done via observation instead of supervised learning.

Structure of the Dissertation

The remainder of this dissertation is organised as follows.

In Chapter 2, goes through a detailed literature review, including traditional anomaly detection techniques, machine learning techniques, Isolation Forest algorithm, and real-time market analysis techniques for perpetual futures market.

In Chapter 3 the research methodology will be elaborated, including the design science approach, the data source and data collection method, the feature engineering pipeline, and the progress of this study regarding anomaly detection.

Chapter 4 presents system design and architecture. It explains all the layers of the system from data collection to user information output and describes the reasoning behind the design decision for module algorithm implementation and probability value aggregation.

Chapter 5 is dedicated to the description of the system implementation which involves the development environment, data processing pipeline, model implementation, system integration and technical hurdles experienced.

In Chapter 6 results are presented based on the two-week observation, model performance analysis, anomaly case studies, comparison of unsupervised model with supervised thresholds, system performance in real-time.

In Chapter 7 will be discussed the results' evaluation according to the research questions, the benefits and drawbacks of the methodology as well as the comparison with the existing academic and commercial systems.

In Chapter 8, summary of findings, contributions, limitations and future work directions are detailed.

Literature Review

Market Anomaly Detection

Financial and digital markets have become increasingly complex and dynamic, driven by high-frequency trading, global interconnectivity, and continuous data generation. In such environments, the identification of market anomalies plays a critical role in understanding abnormal behaviour and potential risks. Market anomalies can be broadly defined as deviations from expected or typical market patterns, often associated with unusual price movements, abnormal trading volumes, or irregular liquidity conditions. Detecting such anomalies is essential for enhancing situational awareness and supporting decision-making in volatile market conditions.

Traditional Monitoring Methods

Early approaches to anomaly detection in financial markets primarily relied on statistical techniques and rule-based systems. These methods typically identify anomalies by detecting deviations from historical averages or predefined thresholds. While such approaches are relatively simple to implement and widely used in practice, they exhibit several limitations. In particular, rule-based systems struggle to capture the complexity and interdependence of modern financial markets. Furthermore, they often fail to respond effectively to rapidly changing conditions, as they rely on static rules and lagging indicators. As a result, traditional methods may generate delayed or inaccurate signals, especially during periods of high volatility.

Machine Learning Approaches

The distinct characteristic of machine learning models compared to rules-based models is that they automatically learn from the input data instead of predefined logic established by humans. This provides greater adaptability to new market events, along with a better detection of the complex relationships between the market variables. Machine learning can be divided into two types, ie supervised learning and unsupervised learning. Supervised learning methods include random forests and gradient-boosted classifiers. These methods require labelled data, or examples of anomalous events with a predetermined outcome, from the past. This is challenging in financial markets, as truly anomalous events are rare, hard to identify and even harder to identify at the time. Furthermore, such events would be easy to identify in the future. On the other hand, unsupervised methods do not require labelled data; these methods model the normal behaviour of the system and use that to identify outliers or major changes. Local Outlier Factor and One-Class SVM are two prevalent unsupervised methods that have been applied to financial data. However, both methods are limited by their high resource use, which limits their real-time applicability (Chandola et al., 2009). Lastly, deep learning methods such as autoencoders and LSTMs (long-short-term memory) networks have also been researched for financial anomaly detection (Goodfellow et al., 2016). These methods are able to learn complex patterns from sequential data, such as market movements over time. However, they would still require large amounts of stable training data, such as what would define normal behavior. In rapidly-changing markets where normal behaviour changes often and rapidly, and there is no objective ground truth, this condition is hard to meet. Overall, it can be seen that the methods that have been developed and researched have pushed towards unsupervised methods to be used. Moreover, these methods are suited for real-time detection of financial anomalies, which is another promising aspect – compared to supervised learning methods. This realization has greatly affected the method used in this study, which will be discussed in detail in Chapter 3.

Isolation Forest and Unsupervised Models

A significant portion of anomaly detection research in financial markets focuses on unsupervised learning methods. Such techniques are important especially in cases, where labelled datasets are unavailable or are very rare, which is the case in most anomaly detection scenarios. Among newer methods, Isolation Forest has gained popularity, since it demonstrates good computational efficiency and effectiveness in finding outlier points in high-dimensional datasets. The algorithm splits the observations using a random splitting , while anomalous points remain much less segmented, therefore requiring less random splits to be identified. Such methods can be generalized to financial datasets where nonlinear relationships are abundant, data might be noisy and may have many dimensions.

Real-Time Market Analysis

Furthermore, the attention has also increased towards the real-time market monitoring apart from the methodological advancement. The streaming data technology aids in minimizing gap between the generation of data and analysis. It facilitates early identification of the observation which deviates from the anticipation. The real-time analytic instruments continuously analyze newly incoming data and help identify the abrupt market changes in a timely manner. However, despite improvements in speed, many existing real-time approaches still have limitations in effectiveness. In particular, they are often not market-specific and fail to consider factors such as liquidity changes, specific trading patterns, or market regime shifts. This limitation reduces their ability to accurately identify significant anomalies in complex financial markets.

Latency is not the only practical implication of real time anomalies detection in financial markets. The field of market microstructures has found that a given price movement may represent an event that is normal on certain volatility regimes, and an anomaly on others. Decisions taken based on static thresholds for finalities of detection will miss this logic and generate a proliferating amount of false positives during trending markets (Golub et al., 2012). – Dynamic methods are preferred in live monitoring environments –Methods that continuously update their reference distribution are more suitable for the application of live monitoring. The perpetual futures markets incorporate price information not available to traditional stock markets. The mechanism of the funding rate, which sends payments from shareholders with long positions to those with short positions, and vice versa, based on the difference between the price of the future and that of the spot, achieves measure of the sentiment leveraged by the market for a perpetual future, number that has no comparative in emerged markets. Alexander et al. (2021) demonstrate in their research that extreme values in funding rates in perpetual future markets of cryptocurrency precede periods of considerable liquidations. The existence of the funding rate allows it to become a leading determinant used by anomaly detection systems that monitor these markets. Information provided by the open interest, which determines the number of outstanding derivative contracts, allows knowing the degree of leverage in the market as a whole. According to Bian et al. (2018), the growing open interest in a directional price movement indicates the entry of leveraged positions in that market, and the reverse movement will generate successive liquidations. The above leads to develop the Liquidation Heatmap Proxy developed in this work, which through the open interest delta allows determining how much leveraged positions there are in the different brackets contemplated in the price.

Another important consideration in anomaly detection is the selection of input data. While traditional approaches have focused primarily on price data, more recent studies highlight the importance of incorporating additional information sources, such as trading volume, liquidity indicators, order-book dynamics, and transaction-level data. The integration of multiple data types can significantly improve the accuracy and robustness of anomaly detection systems.

Research Gap

Despite the growing body of research on AI-based anomaly detection in financial markets, several important gaps remain. First, there is limited work that combines real-time processing with multi-signal analysis, integrating diverse data sources such as price, volume, and liquidity into a unified detection framework. Second, many existing studies rely on proprietary datasets and closed platforms, which restricts reproducibility and limits the ability to benchmark new approaches. Finally, there is a lack of research focusing on open, data-driven systems capable of providing real-time anomaly alerts without relying on predictive modelling.

This thesis aims to address these gaps by proposing an AI-based framework for real-time anomaly detection that integrates multiple market signals and operates on publicly accessible data sources. The focus is on enhancing situational awareness in financial markets through timely identification of abnormal behaviour, rather than on forecasting future price movements.

Research Methodology

Research Design (System-Based Approach)

This is system-based research design also known as constructive or design science (Hevner et al., 2004). Unlike theoretical analysis or application of a technique on an existing dataset, this research produces a functioning or working prototype system. The prototype system is both the research output and for the exploration and exploitation of the researched problem.

The research objectives were met with a suitable design. The study's primary research objectives necessitate a finished system that accepts live data from the actual stock market and process the data to recognize patterns, apply the provided logic to the identified patterns, and yield an actionable output that can be understood and validated. The research objectives seek to answer questions that can only be achieved through a live implementation of a system, and not by discussing the concepts' applicability through a theoretical or abstract viewpoint. Furthermore, if the intention of the research objectives is the attain a working implementation, a prototype-based approach is more justified and reasonable in addressing the research objectives as compared to a simulation- or statistical-based approach.

The analysis framework provided consists of three analysis modules - the price and volume anomaly detection, a liquidation heatmap substitute, and a relative value viewed focus on funding rates analysis. The modules are integrated into an analysis framework consisting of a multi-signal analysis, as discussed in Section 2.6. Lastly, it is essential that each of the modules be independently analyzable, allowing each of the separate dataset's results to influence the multi-signal analysis.

The research will be divided into three main phases, that of developing and implementing the development of the system, the production and operation of the system in which observations and data is either directly correlated to faced questions, or measurements related to the system development is collected and implemented, and, lastly, the evaluation of said data and results in contrast to the questions previously asked. This follows the model of the systematic execution of design science (Peffers et al., 2007), wherein said artefact is developed, implemented and evaluated several times over.

Data Sources and Collection

Data used in this research is assembled exclusively from publicly available market data sourced from the official REST API of the Binance Futures exchange. The use of the Binance exchange for this study is due to the fact that the exchange is the leading platform for trading perpetual futures contracts, accounting for more than 50% of the daily volume of cryptocurrency derivatives trading (CoinMarketCap, 2024). As such, the Binance exchange is the most representative source of data for a cryptocurrency derivatives study. Binance's public API also allows for the extraction of real-time data from the market with no need for valid authentication, ensuring that this research is fully reproducible, which is a vital quality in academic research.

The components: open interest delta, long/short buildup estimation, liquidation level via leverage estimation, liquidity levels estimation are the inventions and concepts used for the creation of the Liquidation Heatmap Proxy. They were the basis of the open-source indicator Liquidation Heatmap Proxy by victhoreb in TradingView Pine Script. The module was designed, developed and implemented from scratch in Python. The Binance Futures REST API was used, and the module was incorporated into the anomaly detection system with the other modules included in this paper.

Three distinct data streams are collected:

The output of the Price/Volume Module will take 1-minute OHLCV data with timestamps. The records will consist of OHLCV, total number of volume, number of volume bought from buyer and seller, and the order flow features as detailed in Section 3.3.

The Liquidation Heatmap Proxy is generated with 5-minute OHLC and Open Interest. The Open Interest on the market - which represents the aggregative market value of future contracts that are open at a certain point of time - will be stored as a measure of leveraged sentiment at the various prices to the estimates on which price levels would reach liquidation on the most utilized leverage (done by Bian et al. 2018). Coinglass provides both software and an API to monetize its own liquidation heatmap, but the data being sourced from their proprietary source, which is a paid closed-source API, user identification cannot be achieved. Without them, the open-interest and prices can be analyzed to measure the sizable and utilized sentiment leveraged on the market and guess on the respective liquidation levels for the leading levels of leverage used.

Funding rate data shows the current perpetual futures funding rate, as well as short historical data. Perpetual futures contracts have a funding mechanism whereby one side of the market needs to pay the other at intervals based on the price difference of the futures and the spot prices. Extreme funding rates indicate excessive leverage on one side of the market and a heightened risk of forced liquidations or short squeezes (Bian et al., 2018).

The rationale for choosing the funding rate over other sentiments is as follows. The Crypto Fear & Greed Index updates once per day and is thus not suitable for real-time analysis. Binance's Long/Short Ratio counts the number of trading accounts instead of contract volume, making it less sensitive to sudden significant trades. Monitoring whale transactions on-chain through data analysis at the blockchain level was considered and discarded as it has no direct correlation to the performance of the futures market, while the sample size is not significant enough to accurately describe general positioning in the market.

Data Preprocessing and Feature Engineering

The raw price and volume data is manipulated to create derived features that aim to capture various market characteristics. A total of nine features are generated for each one minute candle.

Price-based features are features that let us know how the price moved in each of the candles; the percentage return between consecutive candles, the size of the candle in relation to the closing price, and position of the closing price within the candle indicates whether buying or selling pressure was stronger during the period in question.

Volume-based features are interpreting the unusualness of trading volume: the difference in volume between two consecutive candles; a rolling measure that describes how unusual is the current volume in the context of the most recent history; and the weighted measure of how much of the volume was by buyers compared to sellers.

The ±0.04 OFI threshold was derived by trial and error during live operation of the system. A 0.04 OFI ratio indicates that buyers constitute only about 52% of total volume versus only 48% for sellers, which is a minor but persistent skew over 100 candles. Lower thresholds generated excessive noise of misleading signals under balanced market conditions. Higher thresholds produced useful signals too infrequently for monitoring. Although there is no training dataset for supervised optimisation, the signal of 0.04 strikes a useful compromise of sensitivity and specificity. This is a characteristic of the design science approach taken in this study, where the values of parameters are chosen from system operation and observation rather than through supervised optimisation (Peffers et al., 2007). The OFI volume aggregation of 100 candles corresponds to about 100 minutes trading time. This is a time-frame was chosen to aggregate many individual candle signals while still reflecting the current market regime. A 50 candle window was assessed in an earlier prototype, but led to an excessive number of direction changes that did not reflect observable market changes. A 200 candle window was too slow, as it continued to indicate bias toward buying while the market was otherwise showing selling pressure.

The Order Flow Imbalance (OFI) signal is calculated as the net order flow difference between the buyers and the sellers divided by the total volume over a temporal window of last 100 candles. It reflects whether buyers are more aggressive than the sellers (in that case, indicator is positive) or vice versa. The directional signal for trade entry was based on OFI rather than commonly used technical indicators like RSI or MACD as the latter ones are derived from price close and hence are lagging indicators . OFI is derived from the current candle market activity as it reflects how much the buyers are aggressive vs the sellers on the candle execution. The complex math models show that OFI closely related the market price movements in financial instruments over short time horizons(Cont, Kukanov & Stoikov, 2014)

Volatility is calculated as a rolling standard deviation of the price returns. In order to ensure that the current volatility class is classified from a reference that is not the current period, the standard deviation is calculated based on previous candles. A more meaningful classification is returned, with three volatilities classes: NORMAL, ELEVATED and ANOMALOUS .

In the end, momentum features at short and medium horizons express continuation of the price tendency over the recent past and add a trend component to the feature set.

Anomaly Detection Method (Isolation Forest)

The Price/Volume Module has implemented Isolation Forest algorithm (Liu, Ting & Zhou, 2008) to detect anomalies in nine features discussed above. The Isolation Forest algorithm constructs a large number of decision trees. Each decision tree partitions randomly the entire dataset, in the partitions; if the observation is isolated quickly, that means it takes few steps to reach the observation instead of other observations. Based on this decision, if the observation has been partitioned with minimal steps means, it is an anomaly because these entire anomalies will occupy sparse and unusual spaces in data. With this framework, the Isolation Forest identified rare events very efficiently and it does not to know what such an event would look.

The model is configured with 200 estimators and a contamination parameter of 0.05. The number of estimators was selected by the recommendations of Liu, Ting and Zhou (2008), where the authors state that the ensemble performance curve reaches stability above approximately 100 trees; the choice of 200 ensures a conservative margin over this point without introducing computational complexity at each refresh cycle. During the implementation phase, the empirical tests showed that the nature of the distributions of the anomaly scores returned with this configuration were stable between executions. the contamination parameter was set to 0.05, and reflects one of the assumptions postulated before analysis, namely that around 5% of the one-minute candles in any sliding window of 500 candles correspond to truly anomalous observations. Setting the contamination parameter too low will cause the model to ignore genuine anomalous observations, as the assumption with which it operates constrain too much the area in which behaviour can be considered as normal; setting it to high, will cause the model to find anomalies where there are only normal fluctuations. The choice of 0.05 as contamination parameter is common practice in the literature of anomaly detection in a financial context, and was kept as implementation default as a sensible first approach with the understanding that it can be adjusted for specific scenarios with calibration data gathered on that domain. The model's performance will be strongly affected by the adjustments of this parameters when deployed for real-time processing of market data.

The choice of Isolation Forest over other unsupervised anomaly detection techniques was based on the following factors.
The local outlier factor is another method developed by Breunig et al. (2000) based on the evaluation of each observation against its closest data neighbours. However, this is not ideal for real time implementation, as the method becomes too computationally intensive too quickly.
One-Class SVM (Schölkopf et al. 2001) identifies the borderline of normal behaviour in the transformed feature space and is both potentially capable of providing the same level of performance and vulnerable to similar shortcomings when applied to large datasets. Furthermore, One-Class SVM is sensitive to the internal parameters and it is complicated to auto-tune them in an evolving system.
Recently, deep learning techniques such as autoencoders or LSTMs have been utilized for financial anomaly detection (Goodfellow et al., 2016). However, these approaches require either labelled training instances or a baseline history record long enough to determine normal behaviour. This is rarely possible in cryptocurrency markets where conditions constantly change and baseline durations are never certain, while no ground truth is available for anomaly labels.
The use of rolling or standard deviations with simple statistical thresholds is known for its simplicity. Unfortunately, their static nature makes them incapable of adapting to non-stationary regimes. For instance, following a long period of high volatility, a threshold-based approach would produce repeated false alerts rather than signalling an event that significantly deviates from the norm.

Isolation Forest overcomes these limitations. It is unsupervised and does not require labelled data, it remains computationally efficient as the volume of data increases, and it is able to adapt to shifting market conditions as the model is retrained with every cycle of data refresh.

System Design and Architecture

System Architecture Overview

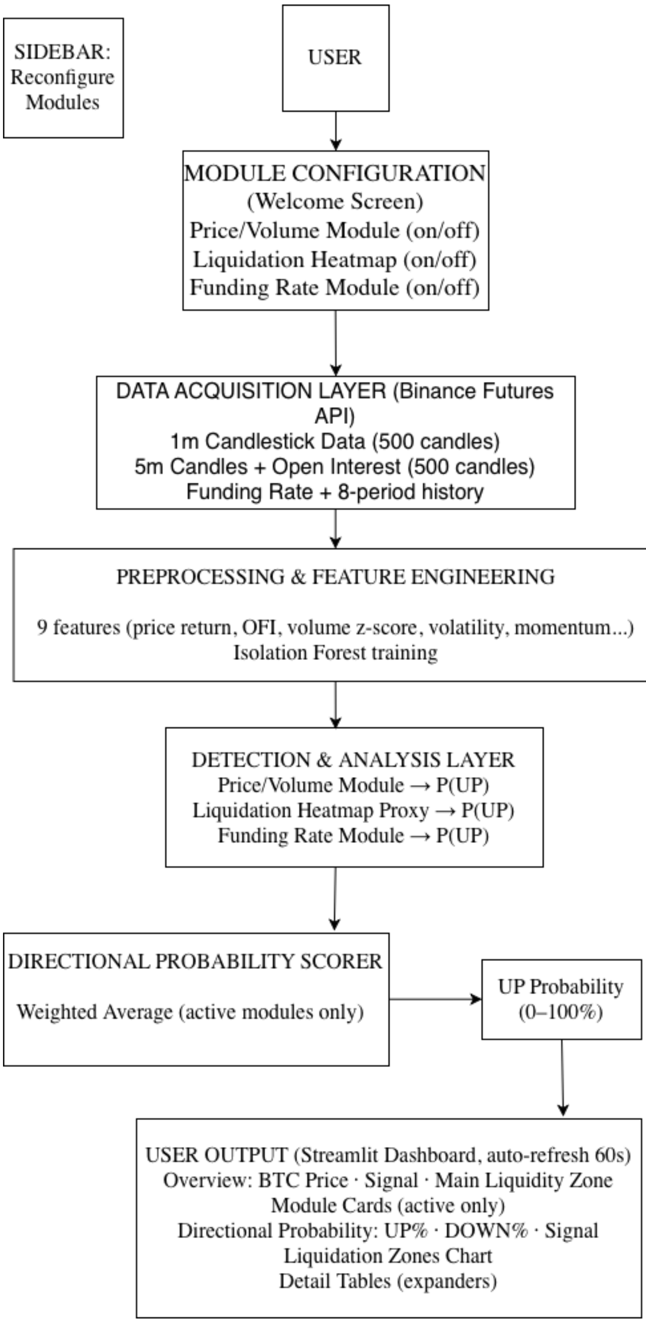

Figure 1. System architecture overview of the anomaly detection framework.

As illustrated in **Figure** "**1**", user — the analyst or researcher who interacts with the system through a web browser. The user selects which analytical modules to activate before monitoring begins and can reconfigure this selection at any point via the sidebar.

Module Configuration — a welcome screen displayed on first launch, allowing the user to enable or disable each of the three detection modules independently. The selection persists throughout the session and determines which modules contribute to the directional probability calculation.

Data Acquisition Layer — There are a number of connection open to the Binance Futures public REST API and allows the programmer to receive three streams of data in parallel: one-minute candlestick price and volumes to study market behavior, five-minute candlestick and open interest to compute the range of liquidations, and funding rate data to assess trader positioning.

Preprocessing and Feature Engineering —captures the update of features that are extracted from raw market data. Features include Order Flow Imbalance, rolling volatility, volume z-score, and momentum features. Isolation Forest model is built on the features in every refresh cycle.

Detection and Analysis Layer — The data routed to three autonomous modules, each returns UP probability (between 0%-100%) - Price/Volume Module (Isolation Forest + OFI), Liquidation Heatmap Proxy, and the Funding Rate Module.

Directional Probability Scorer — a weighted average of the active module probabilities is computed and assigned a UP, DOWN, or NEUTRAL outcome.

User Output — The Streamlit dashboard should show the current BTC price, signal, UP/DOWN probabilities, per-module detail cards, liquidation zones plot, and raw data tables (with expandable sections). The dashboard will automatically refresh every 1 minute.

Data Pipeline (Acquisition, Processing, Streaming)

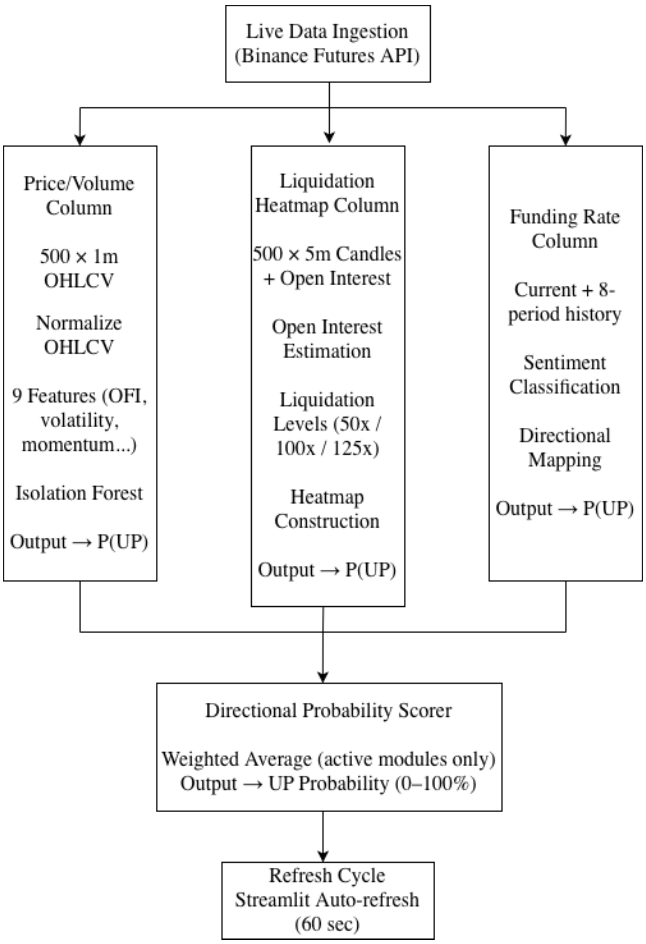

Figure 2. Real-time data processing pipeline across three parallel module columns.

The data pipeline is represented in **Figure 2**.

The detection modules have three respective processing columns that reside in parallel to one another, subsequently converting to the scoring system.

Price/Volume - A connection with the Binance Futures API is created to extract 500 minute candlestick values of BTCUSDT. With the help of normalising raw OHLCV values, nine base features are created. With the Isolation Forest algorithm, the outlier candle is detected for each cycle and its corresponding scores are generated.

Liquidation Heatmap column – The system pulls 500 five-minute candlestick data alongside the current state of open interest. The open interests are compared in following intervals and significant divergence in values suggests changes in the leveraged exposure. Important liquidation levels for three average leverage ratios of 50×, 100×, and 125× are discovered and forms the basis for the corresponding bins as displayed on the heatmap.

Funding Rate column - retrieves the current perpetual futures funding rate and an eight-period historical window. The rate is classified into one of five sentiment categories and mapped to a directional probability.

Directional Probability Scorer receives the UP probability output from each active module and computes the weighted average. Inactive modules are excluded and their weights are redistributed automatically.

Refresh cycle - the entire pipeline is re-executed every 60 seconds via the Streamlit autorefresh mechanism, ensuring that all displayed values reflect the most recent available market data.

Anomaly Detection and Analysis Modules

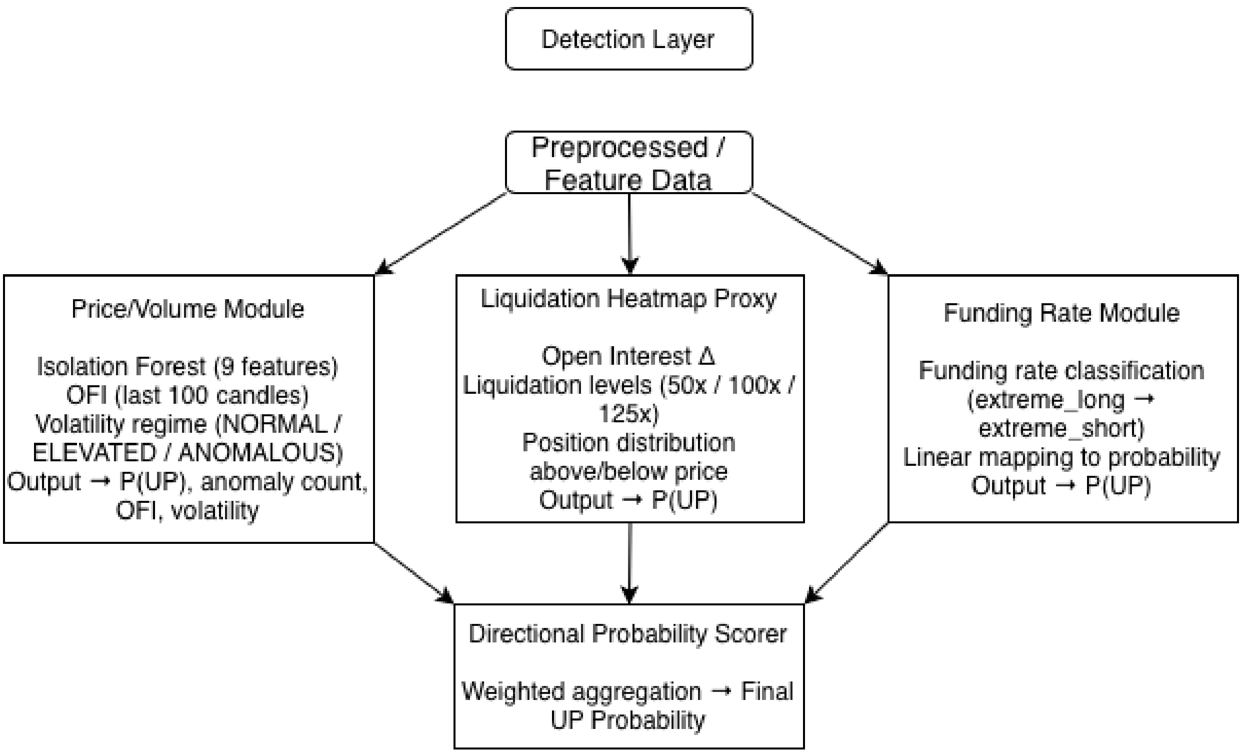

Figure 3. Detection layer showing the three independent analytical modules.

The three detection modules are shown in **Figure 3**. Each module operates independently on its own data stream and produces a self-contained output passed to the scoring layer.

Price/Volume Module — Isolation Forest Algorithm detects pocket of candles that deviate from the recent market dynamics based on the nine engineered features created from the one-minute candlestick dataset. The model provides the direction of the anomaly based on the Order Flow Imbalance ratio derived from the last 100 candles, positive means it's biased to the buying pressure while negatives are net selling pressure. Depending on our knowledge of volatility state (NORMAL, ELEVATED, ANOMALOUS) we weight the anomaly results accordingly. The module returns UP probability, anomaly count, OFI ratio, and volatility state.

Liquidation Heatmap Proxy —The second module estimates the open interest delta, a measure of the number of leveraged contracts concentrated at price levels. The model uses typical liquidation thresholds for a given leverage and predicts where the liquidation will fall for a given concentration of contracts above and below the current price. The ratio of estimated contracts above and the current price is translated to a UP probability, with a higher ratio of contracts above implying greater liquidation demand to the downside.

Funding Rate Module —It pulls the current perpetual futures funding rate and assesses the market sentiment as extremelong, bullish, neutral, bearish, or extremeshort. It linearly maps the funding rate to an UP probability with a mean of 50%, and applies overrides at extremes – to account for the risk of liquidation cascade or short squeeze.

This probability represents how likely the market is to move upward according to that particular module's data. The Price/Volume Module uses the OFI ratio calculated over the last 100 candles as its primary input. Starting from a neutral baseline of 50%, the probability shifts upward when buyers are more aggressive than sellers, and downward when sellers dominate. Each 0.01 unit of OFI contributes a 3 percentage point shift from the neutral midpoint. This base shift is then adjusted according to the current volatility regime: under NORMAL conditions no adjustment is applied, under ELEVATED conditions the shift is amplified by a factor of 1.15, and under ANOMALOUS conditions by a factor of 1.30. The rationale is that order flow signals carry more weight during periods of unusual market activity, when participants are acting with stronger conviction. The final output is kept within the range of 5% to 95% to prevent extreme values from dominating the final score. The Liquidation Heatmap Module calculates its probability directly from the estimated distribution of leveraged contracts above and below the current price. If a larger share of contracts sits below the price, the probability of upward movement is lower, as there is less short-squeeze pressure to push prices higher. Conversely, a higher concentration of contracts above the price increases the UP probability, as those short positions would be forced to close if the price rises. The Funding Rate Module maps the current funding rate linearly to a probability centred at 50%, with each 0.01% of rate shifting the probability by 5 percentage points in the corresponding direction. When the funding rate reaches extreme levels — above +0.1% or below −0.1% — fixed override values of 30% and 70% are applied respectively. These overrides reflect the documented tendency for extreme leveraged positioning to precede sharp reversals, either through cascading liquidations on the long side or short squeezes on the short side (Bian et al., 2018).

Risk Scoring and alert Mechanism

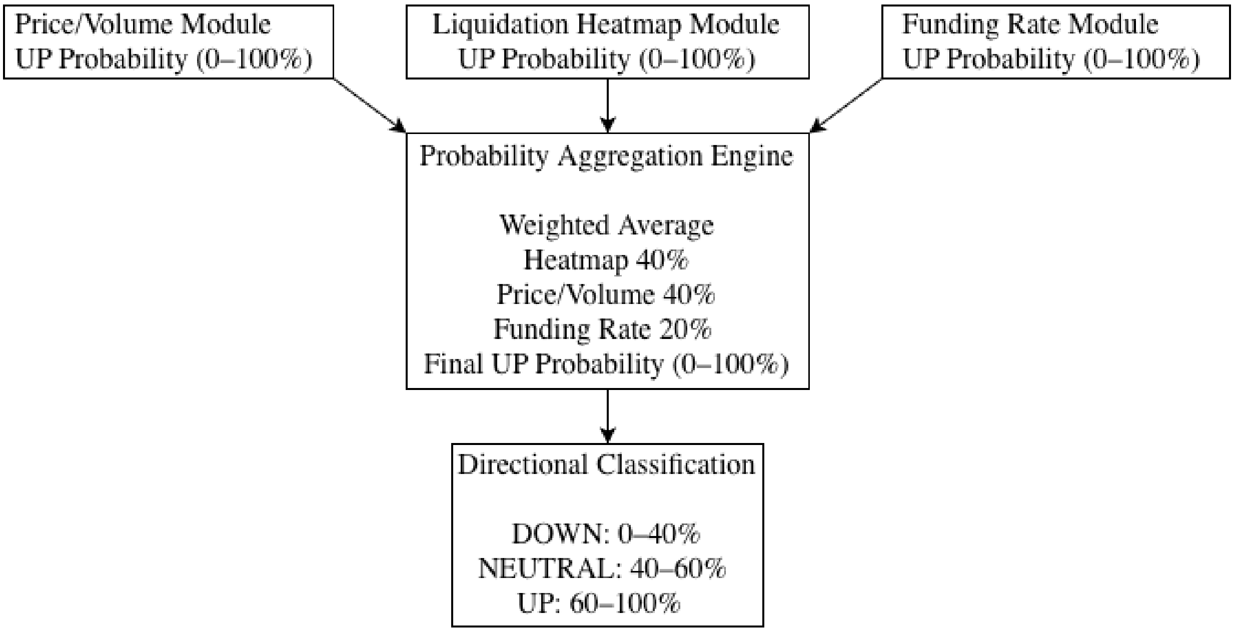

Figure 4. Probability aggregation engine and directional classification thresholds.

As shown in **Figure 4**, a weighted average method is used to merge the outputs of the three underlying detection modules, Price/Volume, Liquidation Heatmap and Funding Rate to generate one output - directional probability.

Each module evaluates the data it works with and gives UP probability between 0 % and 100 %. UP probability is the likelihood that the market will take the upward trend, according to the particular module:

The final probability is computed as a weighted linear combination of the active modules. The default weights are 40% for Heatmap, 40% for Price/Volume and 20% for Funding Rate. If a module is disabled by the user, it will no longer be included and the weight will be automatically redistributed to the active modules.

For these reasons, the Funding Rate is down-weighted to just 20% of the overall importance. Firstly, it is settled only once every 8 hours, meaning what is visible in the meantime is an accumulation of imbalance in a given direction, not a settled price. Secondly, it is the rate for the entire market, not divided into speculation and hedging, meaning it can only be used to have a view on positioning in the market, and is not the live reflected price in the market. Thus, this is deemed to be supporting evidence rather than a primary source of data. The Liquidation Heatmap and the Price/Volume Module are both given 40% because they are current data, refreshed every 60 seconds. Their focus is however very different, with the heatmap telling us the structure of the market in terms of leveraged longs and shorts in relation to the current price, while the Price/Volume Module tells us the live aggression of buyers and sellers in order flow . It would require some form of consistent proof that one is more useful than the other to prioritise one of these. With there having been no way to perform a simulation of this, a fair compromise may be to assign equal importance. This is in accordance with the recommendations of Clemen & Winkler (1999), who conducted research on forecast combination and recommended equal weights as a reasonable choice when no historical and label data is available for optimization.

The decision to use this method of aggregation over any of the alternatives was made in the name of transparency. The previous prototype had used a combination of scores that mapped the final prediction to a probability using a sigmoid function. While mathematically this was a reasonable and viable option, it was necessarily opaque as it would not be possible to determine how each module contributed to the final score – swapping around certain input values could easily produce the same prediction from the model. With the weighted probability average, this concern is eliminated, as the final score is a direct and proportional combination of the score produced by each module in this case, making it straightforward to justify and validate.

Simple ensemble techniques — stacking classifier, neural aggregator, etc. – were not viable the use because they require a labelled historical dataset of correct directional outcomes. There is no right answer for detection of market anomaly in real time, and the notion of right signal cannot be considered.

Distilled directional signal is classified as follows: an UP decision is produced by the probability, which is 60% or more; a DOWN decision is produced if the probability is 40% or less; and anything between these values (40% and 60%) is classified as NEUTRAL in the case of real uncertainty from the market data.

System Workflow

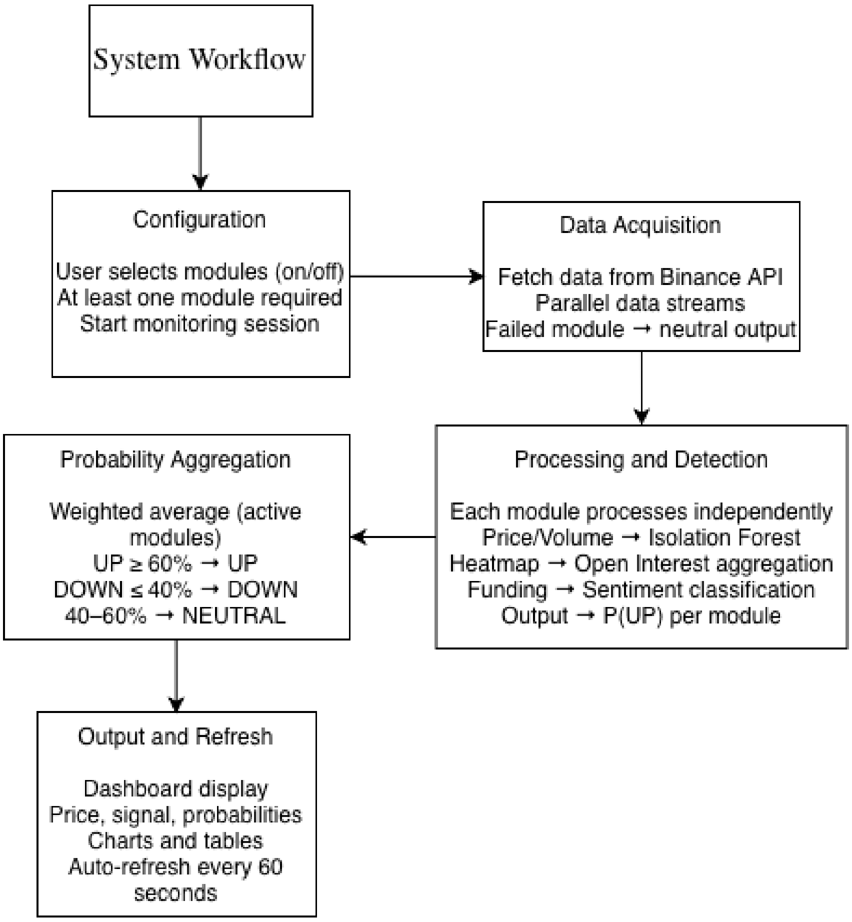

Figure 5. End-to-end system workflow across five stages.

The end-to-end system workflow is presented in **Figure 5**. It proceeds through five stages on each user session and refresh cycle.

Configuration — The user is presented with a module selection screen at first launch. The user has an option to enable or disable any of the three detection modules. The user shall be required to enable at least one module to allow for the proceeding actions. The user can select the confirm button to save the module selection and to begin monitoring.

Data Acquisition — To obtain all three data streams, the unspecified system sends concurrent HTTP requests to the Binance Futures REST API. In the event that one of the requests fails (network error or API unavailable), that specific module returns a neutral default value and the other preforming modules execute normally.

Processing and Detection — each active module processes its data stream independently. The Price/Volume Module engineers features and runs the Isolation Forest model. The Liquidation Heatmap Proxy aggregates open interest changes across price bins. The Funding Rate Module classifies the current rate and historical context into a sentiment category. Each module produces its own UP probability estimate.

Probability Aggregation — the Directional Probability Scorer receives the UP probabilities from all active modules and computes the weighted average. The result determines the directional signal: UP (≥ 60%), DOWN (≤ 40%), or NEUTRAL (between 40% and 60%).

Output and Refresh — The BTC price, signal, UP/DOWN probabilities, detail cards per module, liquidation zones graph, and expandable data tables are showed in the interface. Stages 2–5 will be executed every 60 seconds automatically and all parameters will be updated with the newest data available in the market.

Implementation

Development Environment and Tools

The implementation is done in Python 3.11. It provides all the scientific and analysis libraries which are complete for providing datasets. This package mainly utilizes the following libraries: pandas is used for handling time-series data sets, NumPy for numerical computations, scikit-learn to implement the Isolation Forest algorithm, and requests for the interactions with the Binance REST API.

The user-facing component of the dashboard is based on Streamlit, which is an interactive data application framework that is native to Python. Streamlit was chosen instead of other web frameworks primarily due to pragmatism. Other general web application frameworks, such as Flask or Django, require a separate front-end written in HTML, CSS and JavaScript. Implementing a front-end from scratch for the described dashboard is significantly more work than necessary due to the system's current stage as a research prototype; furthermore, it does not contribute to the research aims of this thesis. A more relevant comparison is Plotly Dash, which is another Python library designed for analytical dashboards. However, Plotly Dash forces a more complex implementation that involves separately defining the layout and callback functions of each component. With Streamlit, it is possible to prototype a fully functioning interactive dashboard in one Python file, with the handling of user interaction and data updating handled automatically on the developer end. This means that the framework is best for quickly prototyping a flexible research dashboard where analytical goals are prioritized over frontend development.

Data is refreshed approximately every 60 seconds, providing near-real time detection and monitoring at all times. Users' first launch of the application will present them with a configuration screen to select which detection modules to enable. The user's selection will persist throughout their session for the application and the scoring system will automatically adapt to reflect only the enabled modules.

Data Integration and Processing

Data streams coming from the three different sources available through the Binance Futures API are processed by separate modules. This separation is afforded by the fact that, although the modules are fed by the same API, each module is concerned with a different data type and a different temporal scale. Thus, failures in any of the modules will not affect the functioning of any of the other two. In the case of the Price/Volume Module, the one-minute raw OHLCV records are passed through the feature engineering pipeline described in Section 3.3. Observe that entire rows corresponding to loss of information due to the initial conditions of rolling computations are removed prior to the feeding of the model. The resulting feature matrix contains around 480 valid rows after accounting the 500 computed candles. The Liquidation Heatmap Proxy module takes the five-minute candles together with the corresponding open interest snapshot, which are matched using a nearest-timestamp join procedure. Processing takes place row-wise on the resulting dataset, estimating levels of liquidation and retaining them in memory. These levels are propagated and updated on each refresh cycle. Levels that have been breached by price during the current cycle are removed from memory and recorded separately, being known as squeezed levels. This group of levels is then interpreted as already liquidated positions. The Funding Rate Module does not require a multi-step process to obtain the required data for its own processing. The current funding rate and historical records are directly extracted from the API and displayed. The current funding rate is processed as a sentiment signal, which is classified using pre-established thresholds. The output from all three modules is returned as dictionaries, which can be considered structured outputs based on a fixed schema. The Directional Probability Scorer receives the output from each module as input. It obtains the requested fields from each of the dictionaries and runs the algorithm to compute the weighted average for each usable drift probability, as described in 4.4.

Model Implementation

The Isolation Forest model is used from the scikit-learn library (Pedregosa et al., 2011). A new instance is created and fit on each refresh using the full feature matrix derived from the current candle window. This was preferable to keeping a state of the model between refreshes, as price dynamics in the cryptocurrency space can change quite radically in a matter of several refreshes. A model fitted on older candle data would not accurately reflect the current volatility baseline. Following the fit, the model calls each candle with a raw anomaly score and label. The raw scores are normalised to a zero-one range (one being the most anomalous observation in the current window). Following this step, the UP probability is derived from the Order Flow Imbalance ratio computed over the last 100 candles. The current OFI ratio and volatility regime is collapsed into a single UP probability prediction, which is passed off to the scorer. The volatility regime is extracted by comparing the current rolling standard deviation of price returns against a historic baseline that executes the most recent 100 candles. This prevents current conditions from bloating their own reference point. Using the mean fit against a baseline, first and second standard deviation bands are extracted, where the former is branded as ELEVATED and the latter as ANOMALOUS.

System Integration and Execution

The entire system is executed as a single Streamlit application file. On start, Streamlit builds the session state and determines if the user has completed the module configuration screen. If not, the application displays the configuration screen, otherwise it proceeds directly to monitoring dashboard. The three detection modules are called in a cached data loading function. The caching here serves the purpose of avoiding repeated API calls and model fitting on every user interaction, as these actions are now done once every refresh cycle. The caching is designed to expire after 55 seconds, which will be the same as the timeout defined in the auto-refreshing command. The auto-refresh cycle is implemented using the streamlit-autorefresh module, which reruns the full application page every x seconds based on the specified parameter. On every rerun, the cached data loading function is called, the output of the modules will be passed to the scorer, and all components in the dashboard screen will be updated accordingly. The entire routine takes about 3 to 8 seconds to complete before the next refresh trigger, depending on the API response and model fitting time, thus providing a generous allowance within the 60-second limit. On the sidebar, a reconfigure button is provided, which will clear the state of the session. This will trigger the welcome screen again during the next rerun. This allows the user to change his module configuration upon user action, without having to restart the application.

Technical Challenges

Several technical challenges were encountered during development.

The public Binance Futures API would sometimes return timeout responses during stress times. This was mitigated by limiting every API call to a timeout of 10 seconds, and by ensuring that every module returns a neutral output by default when its API calls fails, allowing the other modules to continue functioning as is. Liquidation proxy accuracy. The Liquidation Heatmap Proxy calculates liquidation levels using estimates based on open interest changes (growth/decline) and assumed leverage distribution. The positions of each trader on the Binance Futures platform is inaccessible to the proxy, meaning that the heatmap is estimative. The calibration of the proxy depends entirely on whether the premise that changes in open interest means that leveraged positions have been opened on the market with standard levels of leverage usage is correct. This is not guaranteed to always be the case. Streamlit rendering. Each time the user interacts with a component, or the dashboard is automatically refreshed, Streamlit re-executes the whole script. This presented problems when constructing HTML block structures with multiple components, because within different calls of markdown, opening or closing tags would not result in a proper HTML structure. This was overcome by wrapping each one of the dashboard components in a single markdown call that contains the entire HTML block. Isolation Forest retraining latency. Each retraining operation causes a delay of approximately 1-2 seconds, as the model has to be retrained from scratch. During the life of the component, this is acceptable within a 60-second cycle, but could pose a problem if in future versions, the refresh time is set to something significantly lower than this interval.

Results and Evaluation

Experimental Setup

The system was operated for a two-week observation period in real-time using live BTCUSDT perpetual futures data from the Binance cryptocurrency exchange. All three detection modules were enabled for the entirety of the observation period. The display was checked in intermittent sessions at irregular intervals, during which price activity was recorded in various states such as relative calm, rapid movement, and extreme funding rates. No experiment with controlled variables and ground truth dataset labelling was performed, considering that this is consistent with the unsupervised nature of the system and the lack of objective definition for the notion of correctness in real-time predictions triggered by market signalling. Assessment measures revolve around the consistency of signalling among the modules, the reactivity of the entire system to market events, and the consistency of the aggregated probability vis-a-vis the components discovered in the modules. The system was deployed in a local machine with stable Internet service. API requests completed in no more than 1 to 3 seconds under normal circumstances. Isolation Forest model fitting required from 1 to 2 seconds per cycle, which resulted in a refresh latency period of around 4 to 7 seconds.

Model Performance

During normal market conditions, the Isolation Forest model identified 3% to 8% of candles as anomalous, which is consistent with the contamination factor set to 5%. During high volatility situations, characterized by multiple successive price jumps, the percentage of positional anomalies in the last 100 candles increased to 15% to 25%. The volatility regime was labelled ELEVATED or ANOMALOUS for that specific time step. The behaviour of the model during market conditions was stable: during low activity periods, the anomaly score was consistently close the baseline values and the volatility regime was NORMAL. During high activity periods, such as sharp price movements with volume spikes, the model took the reaction to the news in only one refresh cycle. The threshold crossings of the OFI ratio proved to be a responsive directional indicator. The ±0.04 crossings during high activity corresponded extended buying or selling in the 100-candle window. The signal pointed as neutral during balanced order flow periods, confirming the threshold effectively screens the noise resulted from equilibrium situations.

Detected Anomalies and Case Studies

Substantial detection events were observed during the observation period.

Case 1 – Volume spike with neutral price action. One case, where the Isolation Forest returned 12 anomalous candles during the 20-minute period, the volumes traded spiked significantly above the 20-period rolling mean but the price remained range-bound. OFI ratio was near zero indicating that the buying and selling volumes were roughly equal, even though the activity was elevated. Such candlestick configurations can indicate creation of an order book by institutions looking to accumulate or distribute, where they take opposite positions to realize profits but do so discreetly, without affecting the market price.

Case 2 — Extreme Funding rate with corresponding heatmap. The funding rate came out to be 0.12%, which is above the extreme long threshold of 0.1%. The Funding Rate Module returned a P(UP) of 30%, which indicates a high likelihood of liquidation cascade. The Liquidation Heatmap showed a greater number of estimated contracts below the current price, which is consistent with the situation of long holders. Combining the two signals produced a DOWN directional prediction.

Case 3 — ANOMALOUS volatility regime. For three subsequent refresh cycles, the ANOMALOUS volatility regime has been defined during the high price action. With respect to the most recent volume anomaly ON count, it amounts to 22 out of 100 candles. Strong positive OFI ratio equals to +0.11

generates an UP probability of about 83% on the Price/Volume Module. Overall directional signal is UP with 67% probability.

Comparison with Traditional Methods

The detection approach was compared informally to a simple threshold-based monitoring approach. The threshold approach was replicated through fixed percentage thresholding across all trades: an upward anomaly was identified if any candle had a price return above 0.5%, and a volume anomaly was identified where volume was greater than twice the 20-period rolling average.

One major downside to the threshold approach was that it tended to flag far too many false positives during periods of sustained trending. If the market was moving in one direction for a number of hours, every candle that exceeded the threshold would be identified, even though the series of candles were simply continuing the trend, not actually an anomaly. The Isolation Forest, on the other hand, learned and set the anomalous range from the full distribution of recently seen candles, and would not identify the sustained trending moves unless they strayed far from the general pattern.

This finding is consistent with the limitations of static threshold systems identified in the literature review in Section 2.2.

Real-Time System Performance

The system provided consistent real-time functionalities during the testing. The majority of API calls completed successfully. Two instances occurred in which a single API call experienced a time-out due to connection loss; in both situations, the corresponding module returned a neutral default output and all other modules continued to function without disruption.

The 60-second refresh interval is sufficient, changes in conditions will be adequately represented between iterations. The Price/Volume Module utilizes one minute candles, which is compatible with our refresh cycle since it is at least a new tick, providing updates to system relative to market time by at most one minute. Memory footprint remained constant during observation, as levels of liquidations are stored in memory by the storage size parameter, with a non-overhead upper bound.

Discussion

Interpretation of Results

The results of the observation period indicate that this system is able to uncover significant market inefficiencies in real time and communicate these results in an interpretable manner. The answer to the three research questions presented in Section 1.4 is given below.

To answer RQ1, which asked for the methodological differences between AI-based and traditional anomaly detection techniques, it can be said that the comparison presented in section 6.4 proved that the Isolation Forest method is able to learn from the current and recent market behaviour distribution, while threshold-based systems apply a set of fixed rules, no matter the current market situation. That matches the theory explained in the literature review.

In RQ2, we sought to understand the market data properties that were most informative for anomaly detection. The observation period revealed that Order Flow Imbalance and volume z-score features were the most sensitive ones to real market changes. Volume anomalies were frequently leading to or in the same time executes with price change and OFI crossings were embedding directional bias in the next candles.

Through RQ3 we aim to assess how the real-time anomaly notifications could contribute to situational knowledge development about market health. The dashboard presented module-level outputs and aggregate probability estimates in a manner that permits the user to evaluate both the complete signal and the input data's contribution to it. This heuristic serves to foster situational awareness as opposed to mere trust in a single signal.

Advantages and Limitations

Advantages. The advantages of the chosen system architecture is the ability to use public data, allowing the complete reproducibility of the results without access to a proprietary trading platform. Its modular structure allows the user to test system modules independently. The aggregate approach to direct probability allows the user to obtain module-level results that can improve model interpretability in terms of scientific research reproducibility.

Limitations. There are two key limitations associated with this analysis. First, the primary data source – the Liquidation Heatmap – is a noisy proxy. Since access to actual position-level data was not possible, it is unclear how well the derived liquidation areas correspond to true liquidations. Second, as the isolated events cannot be labelled as anomalous or not due to lack of data, the precision of detected anomalies cannot be evaluated using classical classifiers measures of goodness. Finally, the proposed approach features a refresh interval of 60 seconds, which permanently limits the system latency in case of market events.

Positioning within Existing Research

Though the present model is less informative than commercial counterparts like Coinglass, it is more transparent. The heatmaps presented by commercial companies use liquidation data from exchanges which is not available publicly through APIs. These heatmaps are therefore more accurate. However, because the internal liquidation data of exchanges is proprietary, it is impossible to verify the methods used by commercial companies to generate heatmaps. The present model sacrifices accuracy in favor of reproducibility, which is justified in the pursuit of academic research.

With regards the frameworks for research available in literature, the highlighted innovative feature of the solution is its ability to integrate three different sources of complementary information into a single real-time interface. Most of the existing studies focus on one signals only, usually the price or the volume, and dissociate the microstructural indicators from the positioning data. The use of multiple signals directly addresses research gap identified in Section 2.6.

Conclusion

Summary of Findings

The most unanticipated result from this whole endeavor was not the Isolation Forest model per se, but its interaction with Order Flow Imbalance. The discovered candlestick anomaly shapes demonstrated promising results for the model — yet the score alone as not sufficient to determine whether the current signal represented block accumulation by the institutions or merely a random high-volume candle. Only after this data was combined with the Order Flow Imbalance, which provided an intuitive directional prediction for the outcome, was it possible to infer any results. This use of both model score for anomaly detection and directionality information to infer an outcome was completely unforeseen at the design stage, only coming to light from observations during live trading. The identified anomalies were then weighted averaged to a transparent directional probability output shown in a Streamlit dashboard that automatically refreshes the output screen.

The system was successful in its performance throughout the observation period. The anomalous volume activity, extreme funding rates, and high volatility regimes were all correctly identified. An isolation forest was a more robust implementation than a threshold-based system, and the directionality of the order flow imbalance indicator was prompt in reporting signals that aligned with the activity present in the market.

Contributions

This research contributes in following aspects: It presents a repeatable technique to build a multi-signal real-time market irregularity detection system using public sources only. It has direct probability aggregation method where each of the analysis engine generates understandable probability which bypasses the problems of composite score merging, unlike the other methods. The approach generates a proxy for liquidation heatmap which imitates commercial software without unique data access.

Limitations

The first limitation of the system is that the liquidation heatmap is a projection. The second system limitation comes from the directional probability output which is what it is. The parameters calculated by the system describe the prevailing asymmetry of the market: the liquidation zones imbalance or the order flow aggressiveness excess. That is the system is not sensitive to any major news disruptions that are not reflected in the data it analyzes. The Bitcoin halving cycles and other macroeconomic drivers, regulations-related news and rumors, as well as any unexpected changes in the institutional outlook on the market invalidate the simultaneous outputs of all three modules. As a result, the system outputs neutral or contradictory signals not because the price, volume, open interest, and funding rate are at equilibrium, but due to the major driver moving the market. This driver is not in the data and therefore, one cannot gain the visible asymmetry positioning from it. As such, this system is not to be used as a crystal ball predicting the price movement. It is solely revealing the detected microstructural patterns that are currently observed in the market environment depending on which one can make a decision related to the emerged price action.

Future Work

Several avenues for further development are suggested.

First, for liquidations heatmap proxy, order book depth data can be further exploited instead of using the open interest for a certain period prior to current time as it is a more reliable indicator of closeness to immediate liquidity.

Second, the current polling model can be simplified further by creating a WebSocket connection to the exchange instead, thus allowing the application to receive new data as it becomes available rather than waiting for the next poll interval of 60 seconds. This would provide the application with a much faster response to quick market movements.

As a third aspect of development, the model could be generalized to more than one pair of assets. Anomaly detection and correlations between pairs of assets can be very useful for trading. In fact, the divergence of two pairs, which should remain correlated, may be informative, but the current model does not say so.

Fourth, it would be constructible to compare the quality of detection of the system in a structured evaluation study with a set of historical periods that contain well-known market events (for instance, significant liquidation cascades or incidents linked to exchanges).

As previously described in Section 8.3, a potential extension of this work pertains to the usage of additional signals relating to the broader market context. Current implementation is solely reliant on microstructural price action . Events at the macro-level like Bitcoin halving cycles, central banks decisions, regulatory news are currently not accounted for when determining the viability of a directional signal. The system could contain an additional sentiment layer, based on news event classification or on-chain data, that informs whether certain anomalies were driven by forces emanating inside of the market or outside of it. Interpreting the directional signal would benefit from this major contributory factor during periods of market structural break.

Bibliography

Bian, J., He, Z., Shue, K. and Zhou, H. (2018). *Leverage-Induced Fire Sales and Stock Market Crashes*. NBER Working Paper No. 25040. National Bureau of Economic Research.
https://doi.org/10.3386/w25040

Breunig, M.M., Kriegel, H.P., Ng, R.T. and Sander, J. (2000). LOF: Identifying density-based local outliers. *ACM SIGMOD Record*, 29(2), pp.93–104.
https://doi.org/10.1145/342009.335388

Chandola, V., Banerjee, A. and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), pp.1–58.
https://doi.org/10.1145/1541880.1541882

Clemen, R.T. and Winkler, R.L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2), pp.187–203.
https://doi.org/10.1111/j.1539-6924.1999.tb00399.x

CoinMarketCap (2024). *Global cryptocurrency derivatives market share*. [online] Available at: coinmarketcap.com [Accessed May 2026].
https://coinmarketcap.com

Cont, R., Kukanov, A. and Stoikov, S. (2014). The price impact of order book events. *Journal of Financial Econometrics*, 12(1), pp.47–88.
https://doi.org/10.1093/jjfinec/nbt003

Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press.
https://www.deeplearningbook.org

Golub, A., Keane, J. and Poon, S.H. (2012). High frequency trading and mini flash crashes. arXiv preprint arXiv:1211.6667. Available at: https://arxiv.org/abs/1211.6667

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), pp.75–105.
https://doi.org/10.2307/25148625

Liu, F.T., Ting, K.M. and Zhou, Z.H. (2008). Isolation Forest. In: *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM)*, pp.413–422.
https://doi.org/10.1109/ICDM.2008.17

Pedregosa, F., Varoquaux, G., Gramfort, A. et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, pp.2825–2830.
https://jmlr.org/papers/v12/pedregosa11a.html

Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), pp.45–77.
https://doi.org/10.2753/MIS0742-1222240302

Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7), pp.1443–1471.
https://doi.org/10.1162/089976601750264965

victhoreb (n.d.). *Liquidation Heatmap Proxy* [TradingView Pine Script indicator]. Available at: https://www.tradingview.com/script/d2LdGqQO-Liquidation-Heatmap-Proxy-victhoreb/
[Accessed May 2026].

Alexander, C., Deng, J. and Zou, B. (2021). Hedging with Bitcoin Futures: The Effect of Liquidation Loss Aversion and Aggressive Trading. arXiv preprint arXiv:2101.01261.
Available at: https://arxiv.org/abs/2101.01261