Data-Driven Insights in Digital Asset Markets

A Study of Correlation, Anomaly Detection, and Trend Classification Using High-Frequency Price Data

Jesse Palmer
Computer Science
University of Colorado
Boulder, Colorado, United States
jesse.palmeriii@colorado.edu

ABSTRACT

In the fast-moving world of digital asset markets, prices often fluctuate rapidly, sometimes in ways that appear chaotic or unpredictable. This project explores whether short-term price movements contain useful patterns that can be identified using foundational data mining techniques. Specifically, it investigates three central questions: (1) How closely do different cryptocurrency prices move together? (2) Can we detect sudden price spikes that may indicate anomalies? (3) Can short-term price direction be predicted using recent trends?

To explore these questions, we collected high-frequency data from Blocklink Lake, a real-time crypto data API, covering eight major crypto assets. We performed a correlation analysis to examine how asset prices move together, applied a Z-score anomaly detection method to find unusual price movements, and trained a simple Long Short-Term Memory (LSTM) based model to classify short-term trends using recent price data.

My correlation analysis showed that certain asset pairs, such as BTC and ETH or LINK and SOL, moved very closely together, while others like BTC and XRP were less correlated. Interestingly, some assets that might seem unrelated based on their market categorization (e.g., currencies, smart contract token, meme coins) still had moderately strong correlations, showing that market behavior does not always follow intuitive groupings. The anomaly detection system successfully flagged short bursts of unusual price activity, including a sharp spike that reversed a price trend within 30 minutes. My trend classification model reached 63.4% accuracy, demonstrating that even a small set of features from recent price history can be useful in identifying directional tends.

Overall, this work shows that real-time crypto market data carries signals that can be extracted and interpreted using relatively simple tools. These early results highlight the potential of applying data mining techniques to help traders, analysts, and developers make more informed decisions in a dynamic and rapidly evolving market.

INTRODUCTION

The digital asset market moves fast. Prices can swing in seconds, correlations can shift without warning, and trends often appear and disappear in short bursts. In this project, I set out to answer three central questions:

1. How are different cryptocurrencies related in terms of price movement?
2. Can unusual spikes or drops be automatically detected?
3. It is possible to use recent price behavior to predict the short-term direction of a crypto asset?

These questions are important because they touch on real-world challenges that traders, analysts, and researchers face every day. Correlation analysis helps identify which assets tend to move together. This is valuable for portfolio construction, hedging, and understanding how news or market events might ripple through the ecosystem. Anomaly detection can surface sudden, unusual price movements, whether due to news, manipulation, or market stress, and is useful for both real-time monitoring and historical investigation. Finally, short-term prediction offers the potential to anticipate where the market might go next, even if it's only in simple terms like "up" or "down".

To explore these questions, I collected high-frequency pricing data from Blocklink Lake, a real-time digital asset data platform I built to support this kind of analysis. The dataset includes prices from the Coinbase exchange across major assets such as BTC, ETH, SOL, LINK, and others. I used a mix of statistical and machine learning techniques to answer each question: Pearson and Spearman correlations for measuring asset relationships, a Z-score method for detecting anomalies, and a simple LSTM-based neural network for trend classification.

While these methods aren't new, applying them to fresh, high-frequency crypto data, collected every 10 seconds, provided valuable insights. For example, I found that some assets, like Bitcoin and Ethereum, move very closely together, while others behave in less predictable ways. I was also able to detect sharp price spikes and test a working trend classification model that showed real potential, despite being a basic version.

Ultimately, this paper is about insights in noisy, fast-moving crypto data. By using foundational techniques and explaining the steps clearly, I hope to show that even with simple tools, it's possible to uncover meaning patterns in digital markets, and to lay the groundwork for more advanced work in the future.

1 Related Work

This project builds on earlier work that uses data science and machine learning to better understand how cryptocurrency prices behave. Researchers have used similar techniques, such as correlation analysis, anomaly detection, and trend prediction, to study how these digital markets move and change over time.

Correlation Analysis

Looking at how crypto prices move together can help with investment strategies like diversification. One study [1] used Pearson correlation to find which cryptocurrencies tend to move in opposite directions. This helps investors build portfolios that aren't as risky because some assets go up when others go down. Other studies have looked at how these relationships change over time. For example, one paper used a statistical model to show that correlations between Bitcoin and Ethereum can shift depending on the market and news events.

Anomaly Detections

Finding anomalies can help catch problems early or alert traders to opportunities. One study [2] showed how machine learning could spot abnormal activity in blockchain data by analyzing transaction patterns. Another study [3] took a more visual approach using the shape and structure of blockchain networks to find odd patterns that didn't fill the norm.

Trend Classification

Many researchers are also exploring ways to predict whether prices will go up or down in the short term. One study [4] used a type of neural network called LSTM to look at crypto price trends. Their model learned from past price movements and outperformed more traditional tools in predicting what would happen next. More recent research [5] added attention mechanisms, helping models focus on the most important parts of the data to make even better predictions. Combining these techniques with technical indicators like moving averages and momentum has improved the ability to spot trends in Bitcoin prices.

In summary, researchers are using a wide range of tools to better understand digital asset markets. This project adds to the conversation by applying simple but effective statistical methods and machine learning models to real-time, high-frequency crypto data. It helps highlight how techniques like correlation tracking, anomaly detection, and trend prediction can support smarter trading and analysis in this fast-moving space.

2 Dataset

This project used several subsets of high-frequency digital asset price data from Blocklink Lake, a time-series data API which I built and maintain for these purposes, that records prices every 10 seconds from the Coinbase exchange. While the data came from a single source, different analysis sections relied on slightly different windows and asset sets.

For the correlation analysis, I used a 3-day subset of data from April 16-19, 2025, covering eight major cryptocurrencies (BTC, ETH, SOL, LINK, XRP, DOGE, AVAX, SHIB). This dataset provided over 200,000 price points and was nearly complete, with only one gap greater than 30 seconds. These assets were chosen to reflect a mix of well-known and emerging coins, while stable coins such as USDC and USDT since their prices are intentionally stable.

For anomaly detection, I focused on BTC prices over a slightly shifted 3-day window (April 17-20, 2025). BTC was chosen because of its liquidity and consistent price data. This slice included approximately 25,897 records with no gaps or duplicates, making it well-suited for time-sensitive anomaly detection.

For trend classification, I selected a broader 30-day BTC dataset to train and test a machine learning model. This larger window helped ensure enough labeled trends for training. The dataset passed a quality audit, with only a small number of larger gaps (e.g., 700 gaps over 60 seconds out of over 1 million records), which were addressed in preprocessing.

Each data point included three fields:

- **symbol**: the cryptocurrency ticket (e.g., "BTC")
- **price**: the recorded price in USD
- **time**: the timestamp when the price was recorded

This consistent structure enabled a wide range of analysis techniques to be performed, including rolling window detection, pivot table correlation, and supervised model training.

3 Main Techniques Applied

I applied three main techniques to analyze the crypto price data:

- Correlation analysis to understand how asset prices move together
- Anomaly detection to spot unusual price behavior
- Trend classification to predict short-term price directions

Each technique helped explore different aspects of market behavior and answered a unique part of the overall research questions.

3.1 Correlation Analysis

Motivation

Understanding how digital asset prices move in relation to one another is important for identifying diversification opportunities, spotting potential arbitrage setups, or building more sophisticated portfolio strategies. To explore these relationships, I compared how different cryptocurrency prices are correlated over time.

Data Preparation

I used high-frequency price data collected from the Blocklink Lake, which collects digital asset prices every 10 seconds from the Coinbase exchange. This analysis focused on an exact 3-day window, from April 16, 2025, through April 19, 2025. This dataset included eight major cryptocurrencies: BTC, ETH, XRP, SOL, LINK, DOGE, AVAX, and SHIB.

Stablecoins, such as USDC and USDT, were excluded, since their prices are designed to remain constant and do not contribute meaningful insight in correlation analysis.

I first converted the data into a pivot table to make it easier to compare assets. In this format, each row represents a specific timestamp, and each column represents one of the crypto assets. A simplified example is shown below:

Table 1: Sample pivot table showing prices for BTC, ETH, and SOL over aligned time intervals.

Time	BTC (USD)	ETH (USD)	SOL (USD)
2025-04-16 00:00:00	80,000	1,750	125
2025-04-16 00:00:10	80,010	1,745	126
2025-04-16 00:00:20	80,020	1,750	127

This format allows us to line up all the assets across the same time points, so we can compare how they move together over time.

Even though the dataset was very complete, some timestamps were slightly off due to real-world conditions. To fix this and make sure every asset had a value at every time point, I used linear interpolation, which estimates the missing value by averaging the prices before and after it. This works by estimating a missing value based on the prices before and after the gap. For example:

Table 2. Example of a Missing Value in Price Data Before Interpolation

Time	Price
2025-04-16 01:00:00	100
2025-04-16 01:00:10	-
2025-04-16 01:00:20	110

After applying linear interpolation:

Table 3. Data After Linear Interpolation to Fill the Missing Value

Time	Price
2025-04-16 01:00:00	100
2025-04-16 01:00:10	105 (average of 100 and 110)
2025-04-16 01:00:20	110

I also used forward and backward filling, so if there are still missing values, the nearest value from earlier or later in the time series is used.

I ran a simple quality audit to confirm the data quality for this 3-day window. The system recorded 200,564 total records and 200,563 unique timestamps. The average gap between timestamps was 1.26 seconds, with a median gap of just 0.07 seconds. Only one gap exceeded 30 seconds, confirming the data was highly complete and reliable for time-based correlation analysis.

Method

I used two different correlation methods to measure how strongly the asset prices moved together:

- **Pearson correlation** checks how closely two sets of values follow a straight-line pattern. It’s useful when both variables rise or fall together at a constant rate. The caveat is that it can be thrown off by extreme values.
- **Spearman correlation** looks at how consistently two sets of values move in the same direction, even if the rate of change isn’t constant. It ranks the values instead of their actual size, which helps it handle sudden spikes.

Evaluation

The scatterplots of BTC vs ETH, BTC vs XRP, and LINK vs SOL show the general price relationships. In particular, BTC and ETH show a strong positive correlation, meaning their prices tend to move together. The red line on each plot represents the linear trend from the Pearson method.

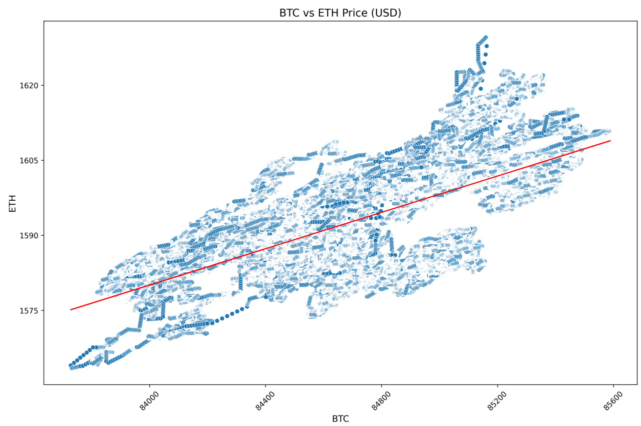

Figure 1: BTC and ETH show a strong positive correlation.

BTC and XRP had the weakest correlation, showing a wider spread in their price relationship.

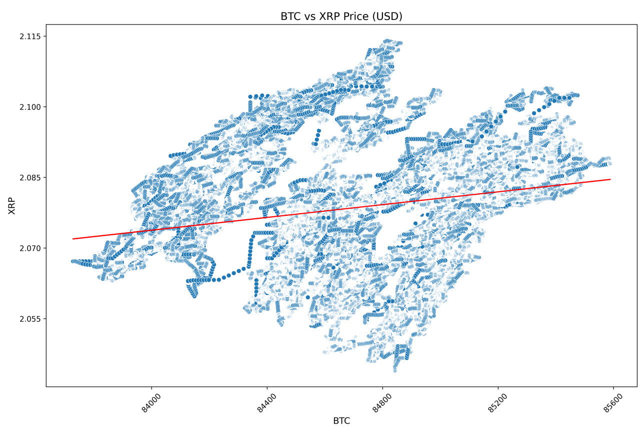

Figure 2: BTC and XRP prices show the weakest.

LINK and SOL had the strongest correlation among all pairs analyzed.

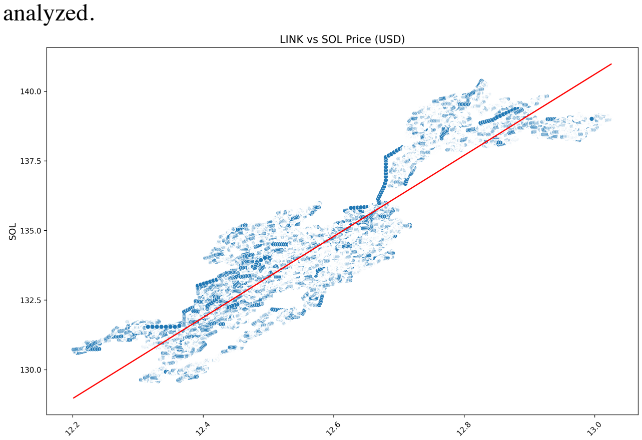

Figure 3: LINK vs SOL show the strongest correlation.

The heatmap below shows the difference between Pearson and Spearman correlations. Red values mean the Pearson correlation was higher, while blue values mean Spearman was higher. The largest gap appeared between AVAX and XRP, where Spearman correlation exceeded the Pearson correlation by 0.17. This suggests their relationship may be non-linear or affected by outliers.

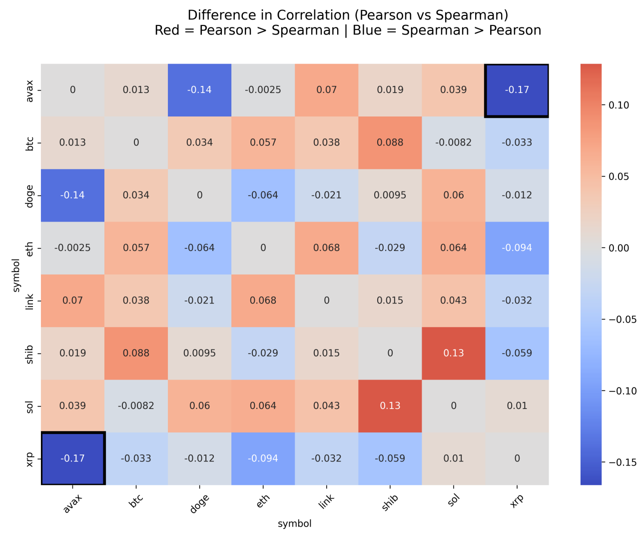

Figure 4: A heatmap showing differences between Pearson and Spearman correlations.

Observations and Implications

- BTC and ETH were strongly correlated, as expected. Their scatterplot showed a clear upward trend.
- LINK and SOL had the strongest correlation among all pairs analyzed.
- BTC and XRP had the weakest correlation, showing a wider spread in their price relationship.
- In general, Spearman correlations tended to be lower than Pearson, except in a few cases, which may hint at subtle non-linear patterns or irregular behaviors.
- Interestingly, assets that belong to very different categories (such as SHIB and BTC) were not necessarily the least correlated. This suggests that price relationships are not always determined by the project type or use case. Some assets may still move together due to shared trading patterns or market sentiment. For example, toward the end of the year, traders may sell their holdings to realize losses for tax purposes which could increase the correlations between assets.

These findings suggest that some assets (like BTC and ETH) often move together, while others may follow more unique or unpredictable patterns. These insights can help traders and analysts decide which pairs to watch for assets might offer portfolio diversification.

Strengths, Limitations, and Future Work

As an informal validation step, I compared the general patterns in my correlation findings against public data from CoinMetrics., a well-regarded digital assets data source. Although their collection methods and time intervals may differ, the results were directionally consistent, suggesting that the core relationships identified here are reasonable.

This correlation analysis provides a simple way to find relationships in digital asset price movements. Using Pearson and Spearman methods allows for comparing simple and more complex patterns in asset prices. For example, Pearson correlation works well when two prices move in a straight-line together, while Spearman can identify patterns where prices move together but not always at a constant rate.

However, the analysis was limited to just three days of price data for the calculations and used one exchange (Coinbase). Future work could expand to:

- Include longer time frames (e.g., 7-day or 30-day windows)
- Compare results across multiple exchanges
- Analyze how correlation patterns shift over time
- Introduce a wide range of digital assets
- Use rolling windows to see trends in correlation strength
- Add features like trading volume, volatility for more context

Overall, this study showed the value of correlation analysis for understanding how digital assets move together and offers a foundation for future, deeper analysis.

3.2 Anomaly Detection

Motivation

Digital assets financial markets are prone to frequent, sudden swings in prices. Factors such as world events, social media posts from influencers, financial news, large purchases, and price manipulation can trigger dramatic fluctuations in the price in the short term.

Being able to quickly and accurately detect these anomalies is essential for traders, data providers, and risk management systems that can assist with past analysis and real-time intervention.

Bitcoin (BTC) was selected in this analysis because it has a long and consistent history of trading data among cryptocurrencies. It's also widely traded across global markets and supported by institutional investors, which helps reduce random price behavior from newer cryptocurrencies that may have smaller market capitalization. These qualities make BTC a strong reference point for testing price anomaly detection methods before applying them to other assets.

Data Preparation and Resampling

Raw BTC/USD price data was collected via Blocklink Lake's high-frequency data platform, specifically its Prices API endpoint. The Blocklink data points that were analyzed were sourced from the Coinbase exchange in 10-second intervals. Over the three days, from April 17th through April 20th, there were 25,897 data points with no gaps or duplicates used for this analysis.

The dataset was resampled to consistent 1-minute intervals using mean aggregations to reduce small amounts of noise that could be caused by irregular trading patterns or brief periods of low activity. While the original dataset had near-perfect continuity, linear interpolation was applied during the resampling to ensure evenly distributed intervals between data points. As a result of this preprocessing, approximately 4,300 consolidated data points were utilized for anomaly detection.

Methodology: Z-score with Median and MAD

To detect anomalies in the dataset, I used a statistical approach based on Z-scores calculated from a rolling median and Median Absolute Deviation (MAD), defined as:

$$Z_t = \frac{0.6745(p_t - median)}{MAD}$$

where p_t is the resampled price at time t. Unlike traditional mean-based Z-scores, this approach is better at handling outliers and skewed data, which are conditions often found in cryptocurrency markets, and avoids being influenced by sharp spikes.

The constant 0.6745 is used to adjust the calculation so that it matches what we would expect from a traditional standard deviation-based Z-score. This adjustment is a standard practice when using the MAD approach. By multiplying by this constant, we make sure that our anomaly thresholds still make sense and stay consistent with traditional Z-score methods.

I used a 60-minute rolling window to define the context for each calculation, and set the anomaly threshold at ±3.5. These parameters were selected after testing various combinations and balancing between detecting meaningful events and avoiding false positives.

Lowering the Z-score threshold makes the model more sensitive, meaning it will detect more possible anomalies, but it also increases the risk of false alarms. Raising the threshold makes the model stricter, reducing noise but possibly missing smaller unusual events. A similar trade-off happens with the rolling window size: a shorter window reacts faster to sudden changes, but can create more false positives, while a longer window gives a less noisy view, but might miss quick shifts. For this project, I chose settings that struck a good balance, detecting important anomalies without creating too much noise.

Anomaly Detection and Cluster Highlighting

Applying the Z-score method to the dataset detected 183 anomalies, comprising 80 upward and 103 downward. Consecutive anomalies spanning over three minutes were grouped into distinct "highlighted clusters," indicating continued volatility or market events during that duration. Nine highlighted clusters were found, which indicates that anomalies tend to happen in short, concentrated bursts rather than isolated incidents.

The largest detected anomaly had a maximum absolute Z-score of 12.7 and occurred at 2025-04-18 21:44 UTC, which coincided with a significant price decrease.

Visualization Integration

To convey these findings, anomalies were visualized using high-resolution SVG plots with the following visual indicators:

- Green circles represent upward anomalies.
- Red circles indicate downward anomalies.
- Orange bands highlight clusters of anomalies over a period of 3 minutes.
- A yellow star represents the single largest anomaly by magnitude.

This visualization allows for quick recognition of isolated spikes and anomaly clusters across the 3-day BTC price series.

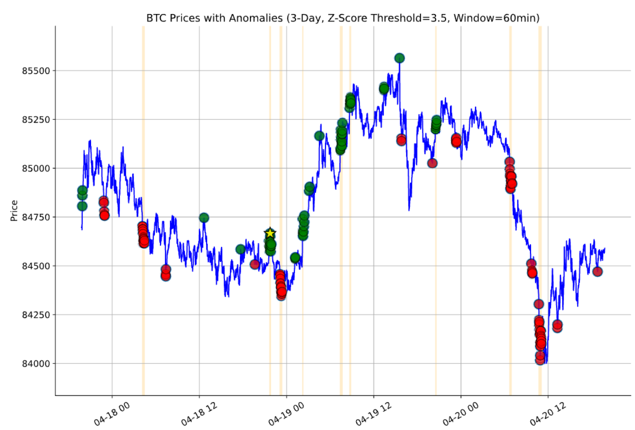

Figure 5: BTC/USD price chart with detected anomalies over a 3-day period, highlighting isolated spikes, directional outliers, and clustered volatility events using Z-score based detection.

Observations and Insights

Anomalies mostly appeared as brief isolated events, yet there were notable clusters that aligned with reversals and directional shifts in pricing. The largest anomaly occurred at a price extreme, followed by a price reversal within 30 minutes, suggesting either a reaction to news or a unusual trading activity. Clusters of anomalies can provide an actionable signal of market shifts, offering valuable insights for short-term risk management and trade execution strategies.

Strengths, Limitations, and Future Work

To spot-check the accuracy of the anomaly detection, I manually reviewed several flagged anomalies on historical price charts. These quick checks confirmed that many of the detected anomalies aligned with real-world sudden price spikes or drops, supporting the reliability of the detection approach used in this analysis.

The adopted Z-score method proved simple, fast, and effective for finding sharp price movements. It is easy to explain and compute, making it well-suited for real-time monitoring or integration with analytics platforms. Because it doesn't rely on complicated models, users can also understand and validate the detected anomalies.

This method works best when market conditions are fairly stable during the analysis window. It may miss slower, more gradual changes that happen over time. Future improvements could include:

- Automatically adjusting the model's sensitivity based on how volatile the market is
- Combining the method with machine learning models that can detect a wider range of unusual market behaviors

For example, a smarter system could learn to spot sudden prices spikes, unexpected crashes, or unusual clusters of trading activity. This would make the anomaly detection system more flexible and better able to adapt to real-world crypto markets.

3.3 Trend Classification

Motivation

Predicting short-term price direction is a common goal in financial analytics, particularly in fast-moving digital asset markets. This section explores whether a simple deep learning model can learn patterns in recent BTC price movements to classify short-term trends as bullish or bearish. While trend prediction can be complex, I focused on building a basic prototype to test feasibility and identify opportunities for future improvements.

Data Preparation

I used 30 days of high-frequency BTC/USD price data from Blocklink Lake, collected at 10-second intervals from the Coinbase exchange. This dataset totaled over 108,000 records. Although there were a few outages in early April, the three-day and seven-day segments were highly complete, with nearly perfect 10-second spacing and minimal gaps.

To make the data suitable for machine learning, I created a few new features:

- **Scaled Price**: Normalized the prices into a [0, 1] range using a MinMaxScaler, so that values stayed within the same scale.

- **Percent Change**: The percentage difference between the current and previous price.
- **Momentum**: The difference between the current price and the price five steps ago.
- **Volatility**: The standard deviation over a short five-point rolling window.

Next, I then labeled each price movement as one of the three categories:

- **Bullish (+1)**: if the next price increased by more than 0.2%.
- **Bearish (-1)**: if the next price decreased by more than 0.2%
- **Neutral**: If the change was between -0.2% and +0.2%, these were discarded to simplify the classification.

A simplified example of how labeling works:

Table 4: Sample Trend Labeling Based on Price Changes

Time	Price (USD)	Percent Change	Assigned Label
2025-04-16 10:00:00	90,000	-	-
2025-04-16 10:00:10	90,060	0.07%	(Neutral, discarded)
2025-04-16 10:00:20	90,180	0.13%	(Neutral, discarded)
2025-04-16 10:00:30	90,300	0.13%	(Neutral, discarded)
2025-04-16 10:00:40	90,500	0.22%	Bullish (+1)

This shows that only meaningful price moves (greater than ±0.2%) were kept for model training.

Sequence Construction

Instead of looking at individual prices, the model was trained to learn from sequences of price movements.

- Each sequence contains 5 consecutive price points.
- The model then tries to predict the trend (Bullish or Bearish) at the next step after the sequence.

Here's clear example:

Table 5: Sample Trend Labeling Based on Price Changes

Sequence Number	Time Steps (Prices)	Next Price	Target Trend Label
1	[90,000, 90,060, 90,180, 90,300, 90,500]	90,600	Bullish (+1)
2	[90,060, 90,180, 90,300, 90,500, 90,600]	90,580	Bearish (-1)
3	[90,180, 90,300, 90,500, 90,600, 90,580]	90.570	Bearish (-1)

Note: In the actual model, prices were scaled between 0 and 1 using MinMaxScaler to help the neural network train more efficiently. However, the examples here use the real USD price values to make the concepts easier to visualize.

- **Sequence 1**: After seeing the prices move up, the next price was even higher (+0.11%), so the model would learn to expect a Bullish move.
- **Sequence 2:** Even though prices were climbing, the next price dropped slightly (-0.02%), so the model labels it as Bearish.

Model Architecture and Training

Once the data was organized into sequences, I built an LSTM model to learn patterns from these price movements. LSTM models are a special type of neural network that are particularly good at working with sequential data, like time series, because they can "remember" important patterns across multiple steps.

For this project, the model was designed to take in sequences of five time steps, each containing for features: the scaled price, the percentage change, the momentum, and the volatility. Based on these inputs, the model was trained to predict whether the next price movement would be bullish or bearish.

The LSTM model was structured into two main LSTM layers, one with 64 units and another with 32 units, each followed by a dropout layer to prevent overfitting. The final output layer used a sigmoid activation function to produce a probability score between 0 and 1, representing the confidence that the trend would be bullish.

The model was trained using binary cross-entropy loss, which is a common way to teach classification models to distinguish between two classes (in this case, bullish or bearish). I tried the model over 20 epochs, meaning it passed through the entire training dataset 20 times, gradually improving its ability to make accurate predictions.

Evaluation and Results

To measure how well the model worked, I used some standard classification metrics:

- **Accuracy**: How often the model's predictions were correct.
- **Precision**: When the model said the market would go up (bullish), how often it was actually right.
- **Recall**: Of all the times market really went up, how many did the model successfully catch.
- **F1-score**: A balanced measure that combines precision and recall into a single number.

The F1-score is especially useful when you care about both types of mistakes (missing real trends and falsely predicting trends) and when the classes aren't perfectly balanced.

The model achieved an overall accuracy of 63.4%. It performed better at predicting bullish trends (70% precision and 67% recall) compared to bearish trends (60% precision and 52% recall). The macro-averaged F1-score was about 0.63, meaning the model found a good balance between catching real trends and avoiding false alarms for both classes.

The confusion matrix below shows the model's prediction breakdown: it correctly predicted 32 bullish and 15 bearish trends, but make 14 false bullish and 16 false bearish predictions.

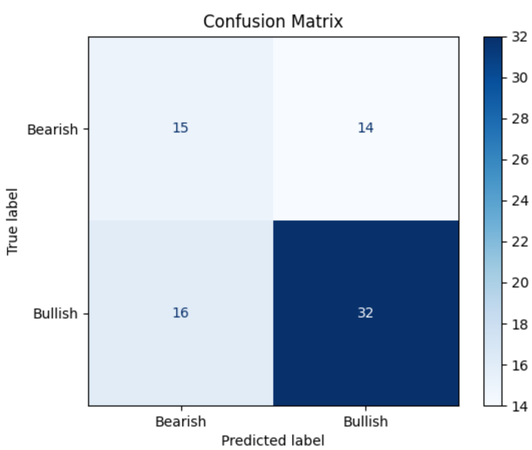

Figure 6: Confusion matrix showing model performance in classifying bullish and bearish trends.

Observations and Implications

Even though this was a basic prototype, the results show that it is possible to predict short-term price direction using on simple price-based features. Because the bullish and bearish examples were fairly evenly split, the model was not heavily biased toward one side.

This type of prototype could be helpful for real-time traders or analysts. For example, a future version could feed into a larger trading system, serve as an early warning signal, or to be combined with other technical indicators to make stronger predictions.

Strengths, Limitations, and Future Work

While no formal external benchmarks were available for validating trend classification, manual inspecting the model's predictions showed they often aligned with visible price patterns. This suggests that the model was able to learn meaningful signals from the data.

One strength of this model is that it was simple, efficient, and easy to understand making it a good fit for real-time systems where speed and clarity are important. There were also some limitations:

- The model only used Bitcoin (BTC) data.
- It relied only on price movements without considering external factors.

To improve performance in the future, it would be valuable to add new types of information, such as:

- Technical indicators (e.g., moving averages or Relative Strength Index)
- Social media activity (e.g., tweet volume, overall sentiment)
- News sentiment (whether new articles are mostly positive or negative)
- On-chain activity (e.g., transaction volumes or wallet behavior)

Overall, these results are an encouraging first step. They show that machine learning techniques can recognize real patterns in raw digital asset prices, opening the door to more powerful forecasting and trading tools in the future.

4 Key Results

This project explored how digital asset prices behave using three techniques: correlation analysis, anomaly detection, and trend classification. Each one helped uncover patterns in how crypto prices move, both on their own and in relation to other coins. Even with just a few days of data and fairly simple tools, we learned a lot about how these assets behave in real time.

Correlation Analysis helped me understand how closely the prices of different cryptocurrencies move together. As expected, BTC (Bitcoin) and ETH (Ethereum) had a strong positive correlation, meaning their prices tend to rise and fall at the same time. Surprisingly, the strongest correlation wasn't between BTC and ETH, it was between LINK and SOL. On the other hand, BTC and XRP had the weakest relationship, meaning the price movements were less closely tied.

One interesting find was that coins that seem very different, such as BTC and SHIB (Shiba Inu), which SHIB is more known as a meme coin while BTC is known as a currency, weren't the least correlated. This shows that just because two assets serve different purposes doesn't mean their prices won't move together. These kinds of insights can help investors figure out which assets are good for diversification and which tend to follow the same trends.

Anomaly Detection focused on spotting sudden price spikes in Bitcoin over 3 days. I used Z-score, which looks for prices far outside the normal range. In total, I found 183 price anomalies, about half were upward spikes and half were downward spikes. While many were just single-point events, some came in clusters, which might point to moments of market volatility or sudden news events. One of the largest anomalies happened right before a sharp price drop, which could help detect short-term market reversals. This kind of analysis could be useful for real-time alerts or reviewing market behavior after the fact.

Trend Classification used the machine learning model, LSTM, to predict whether the price trend over a short window would go up or down. We trained the model using features like momentum and volatility from a 30-day BTC dataset. The model was 63.4% accurate, which in financial markets, even a small edge like this can be valuable when applied consistently. While this model is basic, it proved that recent price behavior contains useful signals that a computer can pick up on. There's clear potential to improve accuracy by adding more data or advanced features.

Together, these techniques showed that digital asset prices do follow meaningful patterns. Correlation analysis showed how assets move together, anomaly detection helped highlight unusual activity, and trend classification explored short-term forecasting. This work provides a starting point for deeper analysis using more advanced methods in the future.

5 Applications

The insights gained from this project can be applied across several areas of digital asset analysis, from day-to-day trading decisions to longer-term portfolio planning and system design.

1. Monitoring and Alerts:

The anomaly detection technique provides a lightweight, interpretable way to identify unusual price movements in real time. This could be used to power alert systems for traders or analysts, helping them respond quickly to sudden market shifts, potential manipulations, or news-driven volatility.

2. Portfolio Management and Diversification:

The correlation analysis revealed which assets tend to move together and which do not. This information is valuable for building portfolios that balance risk and reduce exposure to overlapping movements. For example, know that BTC and ETH tend to be highly correlated suggests they may not offer much diversification when held together.

3. Trend Detection for Short-Term Strategies:

The trend classification model showed early potential for predicting short-term price direction using recent market behavior. This kind of approach could evolve into a tool that informs entry and exit points for short-term trades, especially when combined with other technical indicators or market signals.

4. Foundation for More Advanced Models:

Each of the three techniques, correlation, anomaly detection, and trend classification, proved how simple, well-understood methods can offer meaningful insights. These approaches can be extended into more sophisticated pipelines, such as machine learning systems that can utilized numerous features or a decision-making APIs that support institutional trading or crypto analytics platforms.

5. Educational and Research Use:

Because all three methods were implemented using common libraries and open frameworks, the project serves as a hands-on example for students or developers interested in applying data mining to financial markets. The codebase and structure can be used as a teaching tool or as the foundation for further experimentation.

6 Visualization

To support the research in this project, I created visualizations that highlight key findings such as price correlations, anomalies, and trend classifications. While the static charts in this report were generated separately using tools like Matplotlib and Seaborn, I also added new functionality to an existing interactive web-based visualization that ties directly into the Blocklink Lake dataset.

The interactive site, built using Chart.js, allows users to explore real-time and historical digital prices. For this project, I expanded the platform to include a new feature, overlaying detection price anomalies onto charts. This enhancement directly applied the anomaly detection techniques used in this study.

The interactive web-application is available at https://analytics.blocklink.cloud.

This tool provides a useful way to visualize where significant price changes have occurred, allowing users to validate anomalies, spot patterns, and investigate past anomalies to better understand their potential causes.

Future extensions could include:

- Allowing users to adjust the anomaly detection threshold and rolling window directly through the visualization interface.
- Overlaying trend predictions alongside historical prices.
- Visualizing changing correlation strength between asset pairs over time.

REFERENCES

[1] Dong-Hyun Kwon, Jae-Bum Kim, Jin-Soo Heo, Chul-Min Kim, and Young-Hak Han. 2019. Time Series Classification of Cryptocurrency Price Trend Based on Recurrent LSTM Neural Network. In Journal of Information Processing Systems, 15(3), 555-568. https://jips-k.org/q.jips?cp=pp&pn=680

[2] Swapna Siddamsetti, Chirandas Tejaswi, and Pallavi Maddula. 2024. Anomaly Detection in Blockchain Using Machine Learning. In *Journal of Electrical Systems*, 20(3), 619-634. https://journal.esrgroups.org/jes/article/view/2988

[3] Daniel Ofori-Boaten, Irene Segovia Dominguez, Cagil Akcora, Murat Kantarcioglu, and Yulia R. Gel. 2021. Topological Anomaly Detection in Dynamic Multilayer Blocklink Networks. In Machine Learning and Knowledge Discovery in Databases (pp. 788-803). *Springer*. https://link.springer.com/chapter/10.1007/978-3-030-86486-6_48

[4] Kwon, D.-H., Kim, J.-B., Heo, J.-S., Kim, C.-M., & Han, Y.-H. (2019). Time series classification of cryptocurrency price trend based on a recurrent LSTM neural network. *Journal of Information Processing System*, 15(3), 555-568. https://s3.ap-northeast-2.amazonaws.com/journal-home/journal/jips/fullText/200/jips_680.pdf

[5] Zhang, Y., & Wang, J. (2024). Bitcoin trend prediction with attention-based deep learning models. *Systems*, *12*(11), 498. https://www.mdpi.com/2079-8954/12/11/498