A DATA-DRIVEN QUANTITATIVE FRAMEWORK FOR CRYPTOCURRENCY INVESTMENT PORTFOLIO DESIGN

M.A.N.Maleesha[1], U.P.Liyanage[2]

Abstract

Cryptocurrencies are digital currencies secured by cryptography techniques, and they use blockchain technology to provide decentralised transactions. Bitcoin, Ethereum, and Ripple are examples of over 37 million unique cryptocurrencies today. Often, cryptocurrencies are traded as assets analogous to stock trading, and this study focuses on such trading and investment opportunities. The main objective of this study is to form optimal portfolios by developing strategies based on risk, return, and price classifications. As the first step, feasible currencies were selected using market information and collected hourly price data for three months. After that, based on the volatility spectrum and the market dynamics, price, risk, and return levels were identified. One of the machine learning techniques, namely K-means clustering, is deployed in this level of identification. Concerning the total portfolio investment, small-scale investments ranging from 0$ - 50$, 50$ - 150$, and 150$ - 500$ are prioritised in the analysis. Addressing the different risk aversion behaviours of investors, strategies such as High Risk-High return, Moderate Risk-Moderate return, Low Risk-High return, and Low Risk-Moderate return have been considered in portfolio formation. A Monte-Carlo simulation framework is formed to investigate the performance of the portfolios consisting of cryptocurrencies belonging to different classes of price, risk, and return levels, to address the above investors' strategies. The Markowitz optimisation procedure optimises each of the portfolios. On average, the return of the resulting optimum portfolios exceeded the return of random portfolios that have not been formed according to such a quantitative framework. In conclusion, the steps involving price, risk, and currency level identification, strategically forming portfolios, and selecting the optimal portfolio to be exercised, have formed a rigid quantitative framework that can be utilised in cryptocurrency investment to result in better returns. This framework may be used in forming portfolios of any investment level. However, the machine learning algorithms used in the framework must be recalibrated to the cryptocurrency market changes reflected by the price fluctuation as time progresses.

Keywords: Cryptocurrency, k-means search, Markowitz optimisation, Monte-Carlo simulations, quantitative framework, risk classification

---

[1]Department of Statistics & Computer Science, University of Kelaniya, Sri Lanka
Email: nipunimaleesha8@gmail.com*

[2]Department of Mathematics, University of Colombo, Sri Lanka
Email: prabhathuoc@gmail.com

Proceedings of the 16th International Conference on Business and Information - ICBI 2025 © 2025 by The Faculty of Commerce and Management Studies, University of Kelaniya, Sri Lanka is licensed under CC BY-SA 4.0.

DOI:

Introduction

The modern financial markets offer investors a wide range of options, such as stocks, bonds, Treasury bills, gold, and real estate. Further, investing in currencies is also a common practice among investors. In recent years, cryptocurrency adoption has accelerated dramatically around the world. The global crypto ownership rate is estimated at nearly 12-15% as of 2025. At the same time, over 510 million people hold digital assets, and that is nearly a 50% increment from five years ago. Major markets like India, Nigeria, and Vietnam account for significant growth, while Asia-Pacific now leads all regions with 37.6% of global market share. In the United States, over 15% of adults own cryptocurrencies, and 65% of millennials and Gen Z view digital assets as a preferred investment over traditional stocks.(CoinLaw, 2025). From these trends, cryptocurrencies show that they are new complex investment opportunities while addressing edges that traditional strategies often fail to address. Among the cryptocurrencies, Bitcoin takes a major place. It was the first cryptocurrency, and it was invented by an anonymous entity known as Satoshi Nakamoto in 2009 (Nakamoto, 2008).Bitcoin is now defined as 'Digital Gold' because of its market value and limited supply. One may earn cryptocurrencies by purchasing or mining them, in addition to receiving them as payment for goods and services (World bank group, 2018).

This research introduces a comprehensive quantitative framework that integrates price, risk and return classifications to guide portfolio formation, while other existing studies either focus solely on risk-return optimisation or apply traditional portfolio models directly to cryptocurrency data. Some Studies have demonstrated the success of Machine Learning (ML) algorithms such as Support Vector Machines, Random Forests, and K-means clustering in predicting market trends and optimising portfolios (Rajendran et al., 2024). This framework provides a systematic approach to tailor portfolios for different investment levels and risk appetites, particularly for small-scale investors, by combining machine learning-based clustering with Markowitz optimisation and Monte Carlo simulation. This multi-layered integration offers a novel practical contribution to cryptocurrency investment analysis, bridging the gap between theoretical optimisation models and real-world investment decision-making under high volatility.

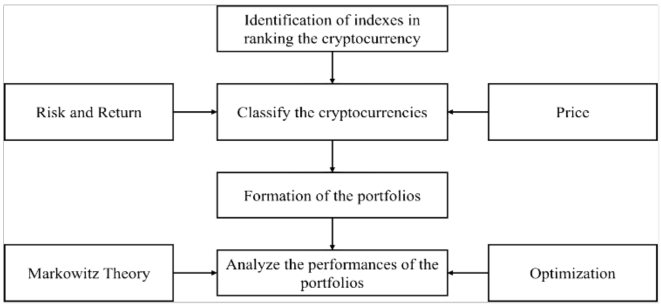
Theoretically, K-means clustering provides a robust mechanism to identify natural groupings of cryptocurrencies based on their statistical properties, reducing heterogeneity within clusters and improving model interpretability(Kocabıyık et al., 2024). The Monte Carlo Simulation method allows portfolio behaviour to be tested across thousands of random market scenarios while enabling stochastic modelling of asset performance under uncertainty (Meher & Mishra, 2024).

Methodology

Often, the cryptocurrency investments are in the range of $ 5000 or below. Nevertheless, the price of cryptocurrency can be as high as $ 60,000 or as low as $ 0.1. Further, investors use reliability ratings to select trustworthy currencies, depending on the financial information available. In this study, the historical hourly pricing data of trustworthy cryptocurrencies with values lower than $ 300 is considered for analysis. Moreover, in the selection of currencies, data consistency is preserved by omitting currencies with missing price values. Therefore, highly priced cryptocurrencies such as Bitcoin and Ethereum were excluded despite their reliability and popularity. Excluding Bitcoin reduces market representation but helps focus on smaller cryptocurrencies with higher growth and diversification opportunities. The currency pricing data is available on many forex platforms. Herein, the 'coinmarketcap', which is a commonly used platform among financial analysts, is used to get the pricing data (Truong et al., 2020). Due to the rapid fluctuations, long-term investments are not practised in cryptocurrencies in the selected price. The dataset consists of the historical hourly prices of cryptocurrencies from November 1, 2024, to January 1, 2024.

The research methodology follows the flow chart shown in Figure 1.

Figure 1
Methodology



(Source: Developed by authors based on literature (2025))

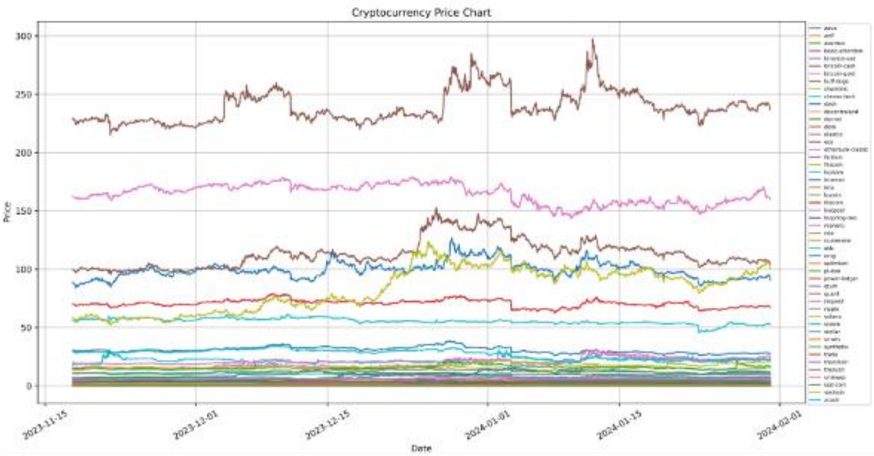
Historical data show that volatility patterns may be reset within a 6-hour time frame, with a 2% margin. Consequently, risk and return patterns should be examined within this same timeframe to capture their changes accurately. Price ranges, on the other hand, exhibit significant changes (a 2% margin) within one week. Accounting for these volatility patterns, risk and return should be classified using price data for 6-hour, 12-hour, and 24-hour periods in determining their associated levels. For price data, the historical data patterns reveal that one week should be considered in classification to obtain its associated levels. To perform these classifications, an appropriate algorithm must be deployed. In this analysis, the K-Means algorithm has provided optimal results, making it the appropriate method for these classifications. K-means was chosen because it efficiently groups cryptocurrencies with similar price and risk patterns, making it ideal for numerical data. The cluster quality was validated using the Silhouette Score for cohesion and the Davies-Bouldin Index for separation.

As the next step, illustrated in Figure 1, portfolios can be formed by selecting cryptocurrencies based on the risk, return, and price levels identified from the classification process. These levels are mapped to common risk diversification strategies: high risk-high return, moderate risk-moderate return, low risk-high return, and low risk-moderate return, allowing investors to choose the strategy according to their risk aversion behaviour. However, in complex investment portfolio formation, numerous risk/return diversification strategies can be employed. However, for small-scale investment portfolios, the chosen diversification is optimal. In forming portfolios, cryptocurrencies belonging to each class are randomly selected based on the strategy undertaken. Each selected set of cryptocurrencies forms a portfolio. This portfolio can be optimised using standard Markowitz portfolio theory. Nevertheless, for performance comparisons, the Monte Carlo technique is employed. This technique generates many such portfolios and compares their expected returns. The computational framework was built using Python (Version 3.11).

Results and Discussion

Adhering to the inclusion criteria defined in the Methodology section, 50 cryptocurrencies have been selected for this analysis. There were nearly 5000 data points because the hourly data were considered. Cryptocurrencies with missing values were excluded since the study emphasises developing a practical investment framework rather than a theoretical model. All data were used ethically in compliance with open-access research standards.

Figure 2
Price profile of selected cryptocurrencies



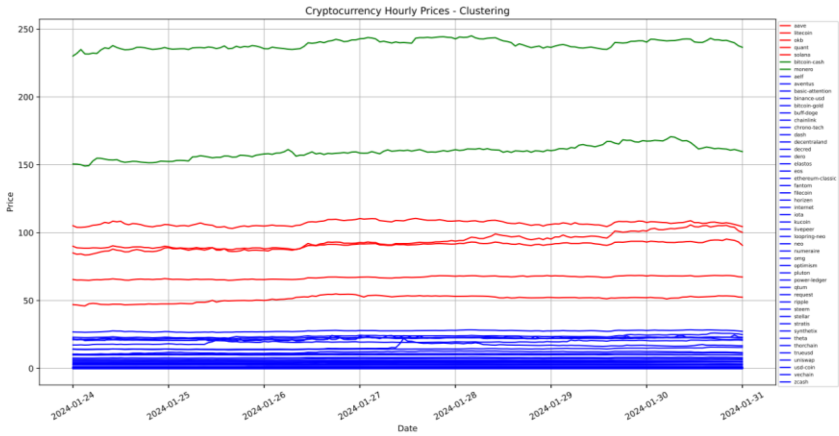
(Source: Developed by authors based on literature (2025))

Figure 2 shows the price patterns of these currencies over the specified period. The figure illustrates the different volatility patterns exhibited by cryptocurrencies. Some show high fluctuations, while others have moderate fluctuations. During the period investigated, the price patterns show relatively stable market behaviour. However, chaotic behaviour is also possible within the cryptocurrency market. Note that most of the currencies (about 80%) have below $ 50 average price. Nevertheless, their volatility is low compared to the other currencies.

Price Classification

In the classification of cryptocurrency prices, one of the machine learning techniques, the K-Means Search, is used. The classification is performed over different time intervals of prices: one week, two weeks, and seven weeks. However, the classification consistently results in three classes, each containing the same set of cryptocurrencies.

Figure 3
Price classification: One-week prices



(Source: Developed by authors based on literature (2025))

Figure 3 shows the price classification considering one week. In this figure, two cryptocurrencies have comparatively high values and are classified into one class. Their price patterns are shown in blue. Five cryptocurrencies fall into a moderate price range, illustrated in red, and form the second class of price levels. The remaining cryptocurrencies, shown in green, have low price levels and form the third class.

The clusters in Figure 3 exhibit some behaviour when considering the 7-week average price as well, due to the market's stability.

Table 1 summarises the price classification along with its respective classes.

Table 1
Price classification, along with respective classes

Price category	Cryptocurrency
Low	self, aventus, basic-attention, binance-usd, bitcoin-gold, buff-doge, chainlink, chrono-tech, dash, decentraland, decred, dero, elastos, eos, ethereum-classic, fantom, filecoin, horizen, internet, iota, kucoin, livepeer, loopring-neo, neo, numeraire, omg, optimism, pluton, power-ledger, qtum, request, ripple, steem, stellar, stratis, synthetix, theta, thorchain, trueusd, uniswap, usd-coin, vechain, zcash
Moderate	aave, litecoin, okb, quant, solana
High	bitcoin-cash, monero

(Source: Developed by authors based on literature (2025))

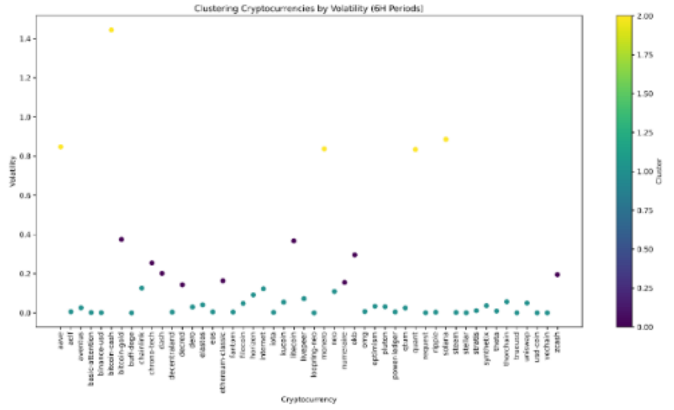
Risk and Return Classification

Risk is typically measured by the volatility of prices over a certain period, while returns are calculated using the following Equation 1.

$$R_t = \frac{P_t - P_0}{P_0} \quad (1)$$

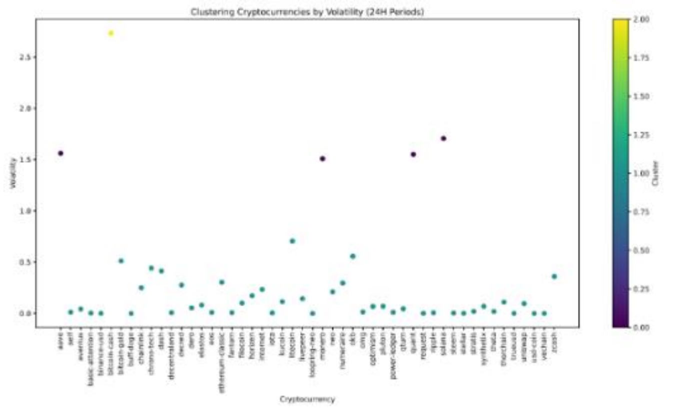
Here, P_0 is the price at the beginning of the period, P_t is the price at the end of the period and R_t is the return over the period. For this analysis, four short-term periods were considered for risk and return calculations: 6 hours, 12 hours, 24 hours, and 48 hours, assuming the short time investments. For these time periods, risk and return levels were also identified using the K-means clustering technique. These time frames were selected to capture the high-frequency volatility characteristics of the cryptocurrency market, where prices fluctuate significantly within short durations Figures 4 and 5 show the resulting classification for the risk levels for the 6-hour and 24-hour periods, respectively.

Figure 4
Risk classification: 6-hour period



(Source: Developed by authors based on literature (2025))

Figure 5
Risk classification: 24-hour period



(Source: Developed by authors based on literature (2025))

As the figures illustrate, there is not much difference among the risk classes, indicating consistency. The results were consistent with the other two time periods as well. Table 2 summarises the risk classification of the currencies along with their respective classes.

Table 2
Risk classification

Risk Class	Cryptocurrency
Low	aelf, aventus, basic-attention, binance-usd, buff-doge, chainlink, dero, elastos, eos, fantom, filecoin, horizen, internet, iota, kucoin, livepeer, loopring-neo, neo, omg, optimism, pluton, power-ledger, qtum, request, ripple, steem, stellar, stratis, synthetix, theta, thorchain, trueusd, uniswap, usd-coin, vechain
Moderate	bitcoin-gold, chrono-tech, dash, decentraland, decred, ethereum-classic, litecoin, numeraire, okb, zcash
High	bitcoin-cash, monero

(Source: Developed by authors based on literature (2025))

As mentioned above, the return classes were also identified. The results are summarised in Table 3 below.

Table 3
Risk classification

Return Class	Cryptocurrency
Low	internet, kucoin, optimism, solana
Moderate	zcash, aelf, aventus, bitcoin-gold, chainlink, chrono-tech, decred, elastos, ethereum-classic, fantom, filecoin, iota, numeraire, okb, synthetix, theta, uniswap, vechain
High	aave, basic-attention, binance-usd, bitcoin-cash, buff-doge, dash, decentraland, dero,eos, horizen, litecoin, livepeer,loopring-neo, monero, neo, omg, pluton, power-ledger, qtum, quant, request, ripple, steem, stellar, stratis, thorchain, trueusd, usd-coin

(Source: Developed by authors based on literature (2025))

Portfolio formation and Optimisation

To construct an optimised portfolio, the allocation strategy was designed based on three key factors: price, risk, and return clusters. These clusters were determined through prior analysis and classified into three distinct categories: low, moderate, and high. The main aim of this algorithm is to address the risk-aversion behaviour of investor profiles. According to the available capital, a dynamic price selection mechanism was implemented.

- Investment Amount: \$10 - \$1000: The portfolio primarily consists of low-priced cryptocurrencies.
- Investment Amount: \$100 - \$10,000: The focus is on moderate-price cryptocurrencies.
- Investment Amount: Above \$10,000: The portfolio allocates a higher weight to high-priced cryptocurrencies.

This step ensured that the portfolios' price structure matched the investor's capital size.

(1) Allocation Strategy: The allocation strategy was guided by defined weights for risk tolerance, price range, and return expectations. Investor's risk aversion behaviour was introduced as conservation, balanced, or aggressive. Then, depending on that risk tolerance, a weight distribution was applied to low, moderate, and high-risk clusters. The strategy was further refined based on expected returns and the price range, which related to the investment amount.

For example, an Aggressive investor with a high return expectation and an investment amount in the low price range would allocate as follows:

- Risk Allocation: 20% to low-risk, 30% to moderate-risk, and 50% to high-risk cryptocurrencies.
- Price Allocation:70% to low-price, 20% to moderate-price, and 10% to high-price cryptocurrencies.
- Return Allocation:10% to low-return, 30% to moderate-return, and 60% to high-return cryptocurrencies.

Cryptocurrencies from each cluster were randomly selected based on the above-weight allocation. From that, the aim was to ensure balance and diversity within the portfolios, and to align the investor's expectations at the same time.

(2) Portfolio Optimisation: While addressing the allocation constraints, the selected cryptocurrencies were then subjected to Markowitz portfolio optimisation, aiming to maximise the Sharpe ratio. This was achieved through the following steps.

- Return Calculation: The percentage change in the historical price data for each selected cryptocurrency was calculated, producing a return dataset.
- Optimisation: Using the return data, portfolio weights were optimised through the Markowitz method, which balances expected returns against portfolio volatility.
- Efficient Frontier Analysis: The optimised portfolios were plotted on the efficient frontier to identify the portfolio with the maximum Sharpe ratio.

The final finding of the portfolio construction was that the allocation aligned with the investor's risk, return, and price preferences while optimising for the highest potential return given the specified constraints. Table 4 indicates the optimum portfolios for the High risk- High return strategy.

Table 4
Optimum portfolios for the High risk- High return strategy

0\$-100\$	100\$-1000\$	1000\$- above
No portfolios available	aave-0.4397 quant-0 monero-0.56029 bitcoin-cash-0	aave- 0.90696 quant - 0.0913
	Return 5.2642e-05	Return 5.3522e-05

(Source: Developed by authors based on literature (2025))

Table 5 indicates the optimum portfolios for the Moderate risk - Moderate return strategy.

Table 5
Optimum portfolios for the Moderate risk - Moderate return strategy

0\$-100\$	100\$-1000\$	1000\$- above
zcash-0.728572 bitcoin-gold-0.10871 thorchain-0.42404	kucoin-0.162535 zcash-0.728572 bitcoin-gold- 0.1025	bitcoin-gold-0.18087 zcach-0.72848 numeraire-0.16277
Return 0.000261	Return 0.0002635	Return 0.00265

(Source: Developed by authors based on literature (2025))

Table 6 indicates the optimum portfolios for the Low Risk-High return strategy.

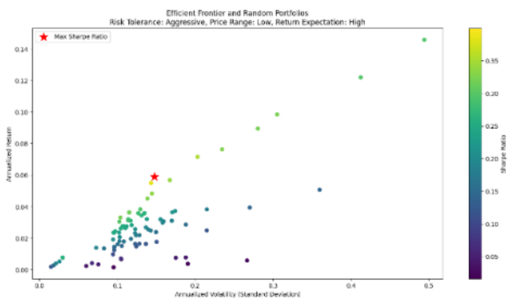
Table 6
Optimum portfolios for the Low Risk-High return strategy

0-100	100-1000	1000$- above
basic-attention-0.015417 binance-usd-0.0435 buff-doge-0.08199 dero-0.1825 loopring-neo-0.050224, thorchain-0.2128 trueusd-0.086 ripple-0.18 stratis-0.05653 usd-coin-0.0372	basic-attention-0.015417 binance-usd-0.0434 buff-doge-0.08199 dero-0.18248 livepeer-0.546 loopring-neo-0.050224 pluton-0.0537 ripple-0.18 thorchain-0.2128 trueusd-0.086 usd-coin-0.0379	basic-attention-0.0095 binance-usd-0.04206 buff-doge-0.0792 dero-0.179 livepeer-0.0545 loopring-neo-0.0486 pluton-0.0489 thorchain-0209 ripple-0.1753 trueusd-0.085 usd-coin-0.0365 aventus – 0.0313
Return 0.0001797	Return 0.0002635	Return 0.00265

(Source: Developed by authors based on literature (2025))

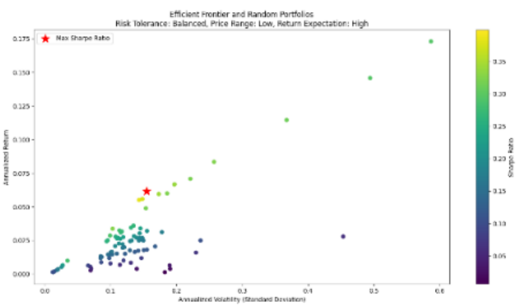
Figure 6 and Figure 7 illustrate the Efficient Frontiers generated from the portfolio optimisation process, showing the potential risk-return trade-offs for different portfolio compositions.

Figure 5
Efficient Frontiers generated from the portfolio optimisation process



(Source: Developed by authors based on literature (2025))

Figure 6
Efficient Frontiers generated from the portfolio optimisation process



(Source: Developed by authors based on literature (2025))

Conclusion

The cryptocurrency market is a volatile market, unlike traditional stock exchanges. And there is no centralised association to govern the transactions. This is very important to investors because they don't have to pay any transaction fees. The main goal of this study is to develop a novel framework to form portfolios in the cryptocurrency market. According to the above methodology, a novel version of the framework was developed. Price, Return, and Risk classifications play major roles in this framework. According to the results, the clusters have high accuracy. And the number of clusters always became three. The levels are defined as High, Moderate, and Low. This framework requires continuous practice with the new prices. If someone follows this sequence of steps, they will have the optimum portfolio. Using this result, the investor can identify the portfolio spectrum as well. Additionally, need to mention that this study is about only for the quantitative basis of the cryptocurrency market. The qualitative measures also affected the return. With proper investigation of the market dynamics and the investor decisions cryptocurrency market has the potential to give continuous returns to the investor. The News and sentiment changes can strongly impact cryptocurrency returns then the framework may be affected. Future studies should integrate AI-based techniques to improve the framework.

References

CoinLaw. (n.d.). *Cryptocurrency adoption statistics 2025*. Retrieved October 13, 2025, from https://coinlaw.io/cryptocurrency-adoption-statistics/

Kocabıyık, T., Karaatlı, M., Özsoy, M., & Özer, M. F. (2024). Cryptocurrency portfolio management: A clustering-based association approach. *Ekonomika, 103*(1), 25–43. https://doi.org/10.15388/Ekon.2024.103.1.2

Meher, P., & Mishra, R. K. (2024). Risk-adjusted portfolio optimization: Monte Carlo simulation and rebalancing. *Australasian Accounting, Business and Finance Journal, 18*(3, Special Issue), 85–101. https://doi.org/10.14453/aabfj.v18i3.06

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. https://bitcoin.org/bitcoin.pdf

Rajendran, H., Kayal, P., & Maiti, M. (2024). Harnessing machine learning for predicting cryptocurrency returns. *Global Business Review*. Advance online publication. https://doi.org/10.1177/09721509241226575

Truong, C., Oudre, L., & Vayatis, N. (2020). A selective review of offline change point detection methods. *Signal Processing, 167*, Article 107299. https://doi.org/10.1016/j.sigpro.2019.107299

World Bank Group. (2018). *Cryptocurrencies and blockchain*. https://doi.org/10.1596/978-1-4648-1292-7