PAPER ID: 20260201004

Cryptocurrency Portfolio Management Using the Yamanaka Phenomenon

Sandeep Bhattacharjee
Assistant Professor, Amity University, Rajarhat, Newtown, Kolkata, West Bengal 700135, India
E-mail: sandeepbitmba@gmail.com

Abstract: This research investigates the role of feature engineering in the analysis and prediction of cryptocurrency markets by employing Yamanaka factors. Through the integration of data preprocessing, engineered feature extraction, and clustering methodologies, the study evaluates the influence of these features on market insights and predictive models. Historical cryptocurrency data from Yfinance has been analyzed, with K-Means clustering applied on both (with and without) the inclusion of Yamanaka factors. Performance assessment has been done using the Silhouette Score and the Davies–Bouldin Index. Furthermore, the study also investigates how engineered features enhance the precision and robustness of cryptocurrency price prediction models. A significant enhancement in clustering and predictive performance was observed upon the application of feature engineering, Which, enabled constructive insights for investment strategies and risk management. This research also highlights the benefit of feature engineering in refining analytical and predictive models within cryptocurrency markets.

Keywords: Cryptocurrency Market Analysis, Davies–Bouldin Index, Feature Engineering, Predictive Modeling, Risk Management, Silhouette Score, Yamanaka Factors

Introduction:

The sustained high impact growth of the cryptocurrency market can be owed to the increasing usage of machine learning (ML) techniques for market analysis, price prediction, and risk assessment. Feature engineering can be regarded as a highly detailed process that may include selection, transformation, and construction of meaningful variables aimed at improving model outcomes (Chong et al., 2017). Many financial factors may remain untouched, because of the inherent volatility and multifaceted nature of digital assets. Subsequently, researchers have been focusing on the application of the concept of Yamanaka Factors (**Oct4, Sox2, Klf4, and c-Myc**), a market driven metrics, for their potential utility in cryptocurrency trading (Yamanaka & Shimizu, 2020). Ré et al. (2014) not only tried extracting meaningful patterns from raw data to create comprehensive and reliable knowledge repositories using constructing knowledge bases but also portrayed key challenges and potential solutions in feature engineering for knowledge bases. Raw market data is often known to lack structured patterns that generate effective predictive analytics. Feature engineering acts as a reconstruction agent that can recreate patterns for predictive analytics. Technical indicators such as moving averages, the relative strength index (RSI), and Bollinger Bands are commonly used in traditional markets to generate actionable insights (Chong et al., 2017). Cryptocurrency markets are known to be associated with additional factors, including market sentiment, blockchain analytics, and liquidity dynamics—exert a profound influence on price movements. A novel approach with the integration of momentum-based and liquidity-driven indicators was one of the earliest applications of Yamanaka factors for equity markets (Yamanaka & Shimizu, 2020).

The future of human activity recognition (HAR) by juxtaposing deep learning techniques with traditional feature engineering methods was explored, where domain-specific feature extraction was compared with unsupervised feature learning in deep neural networks for applications (Kanjilal and Uysal ,2021). Research findings based on evaluation of 64 different learning representations, classifiers, and datasets indicate the competitiveness of feature engineering as compared to deep learning models particularly for large datasets. Another study critically evaluated the effectiveness of Yamanaka factors in cryptocurrency market analysis by synthesizing them into ML models. Zheng and Casari (2018), offered a comprehensive investigation of the role of feature engineering in the machine learning pipeline focused on practical methodologies for extracting and transforming numeric feature representations of raw data that with suitable formats for ML models.

Literature review

1.1 Feature engineering

The studies reviewed collectively underscore the critical role of feature engineering and advanced data processing techniques in enhancing the performance of machine learning models across diverse domains. Khare et al. (2023) provide a systematic review of data fusion techniques for the automated detection of developmental and mental disorders in children, covering nine major disorders and emphasizing the importance of signal analysis, feature engineering, and decision-making models. Similarly, Mahajan et al. (2023) evaluate ensemble learning methods for disease prediction, finding that stacking achieves the highest accuracy, followed by voting, bagging, and boosting. These studies highlight the transformative potential of integrating multiple data sources and advanced learning techniques to improve diagnostic accuracy in healthcare.

Many studies underscore the impact of feature engineering and other advanced data processing techniques across various domains. A systematic review of literature by Khare et al. (2023) discovered signal analysis, feature engineering, and decision-making models for the automatic detection of developmental and mental disorders in children, covering nine major disorders. Similarly, Ensemble learning methods were critically used for disease prediction with stacking method followed by voting, bagging, and boosting (Mahajan et al. ,2023). These studies highlight the transformative potential of integrating multiple data sources and advanced learning techniques to improve diagnostic accuracy in healthcare. Feature engineering has emerged as a central theme across many studies, with researchers mostly supporting its prioritization for enhanced model performance. Verdonck et al. (2024) and Heaton (2016) laid emphasis on the aspects of feature design and data quality, citing empirical evidence for feature selection and transformation. Cognito, an automated feature engineering system that iteratively improves feature construction, signifying its efficacy on real-world datasets, was introduced by Khurana et al. (2016). Kuhn and Johnson (2019) further provided evidence on practical applications of feature engineering and selection strategies. These works collectively promoted the efficacy of feature engineering as an important step to enhance other novel algorithms.

The integration of deep learning into feature engineering has been a matter of continuous focus in many studies with results proving the superiority of feature engineering in extraction and automation. Fan et al. (2019) and Long et al. (2019) identified deep learning based approaches for energy prediction and stock price movements, that showed significant improvements using feature engineering method. Seide et al. (2011) explored the application of feature engineering that inherently learns complex structure in speech recognition. Additionally, Galli (2024) discovered the usage of python programming for feature engineering covering advanced tools and real-world applications. These studies collectively highlight the evolving landscape of feature engineering, where deep learning and automation are increasingly pivotal in achieving state-of-the-art predictive performance across various fields.

1.2 Cryptocurrency volatility

Cryptocurrency markets are mostly characterized by volatility patterns influenced by a combination of macroeconomic factors, financial market dynamics, and internal market mechanisms. Much research has intensively being engaged on disentangling these influences by means of advanced econometric models, focusing on both the transmission of volatility and the development of predictive frameworks. Studies such as those by Yen and Cheng (2021) and Conrad, Custovic, and Ghysels (2018) highlight the significant impact of external macroeconomic factors, such as economic policy uncertainty (EPU) and global economic activity, on cryptocurrency volatility. A higher EPU intensifies volatility during periods of market stress (Yen and Cheng,2021), while a nuanced relationship may exist between Bitcoin's volatility and traditional financial markets, such as the S&P 500 (Conrad et al., 2018). These findings reveal the sensitivity of cryptocurrencies to broader economic forces, despite their decentralized nature, and advocate that macroeconomic indicators actually play a pivotal role in shaping their volatility. The relationship between cryptocurrencies and traditional financial markets in their own ecosystem has been another vital domain of investigation. High volatility spillovers between cryptocurrencies and major financial markets challenged the notion of cryptocurrencies as isolated assets (Liu and Serletis, 2019). Bitcoin's dominant role in influencing other cryptocurrencies such as Ethereum and Ripple was demonstrated by the work of Yi, Xu, and Wang (2018). These subjects indicate the existence of inter volatility within cryptocurrencies and intra volatility with other traditional financial systems. The existence of such dual interconnectedness highlights the need for comprehensive in-depth research of both digital and conventional financial markets absorbing shocks from the financial market.

One of the central challenges has been forecasting cryptocurrency volatility, the key to which remains predictive accuracy. Advanced models, such as the Score-Driven model with Generalized Hyperbolic Skew Student's (GHSKT) innovations, outclasses traditional GARCH models by being more efficient in capturing not only long-memory characteristics but also asymmetric volatility responses (Catania, Grassi and Ravazzolo, 2018). The need for models tailored to specific trading environments has been accrued to strong interdependence and asymmetric responses of cryptocurrencies to market shocks. These studies together highlight the growing need for justifying appropriate volatility tools based on the time horizon and market context and seem to offer valuable tools for investors and policymakers. Overall, the research underscores the multifaceted nature of cryptocurrency volatility, formed by a convergence of external and internal factors, and indicates the ongoing need for innovative modeling approaches to augment market understanding and risk management.

2.3 Feature engineering in cryptocurrency:

Fresh advancements in cryptocurrency research has been spurred by the integration of artificial intelligence, deep learning and blockchain-based methodologies for meeting several challenges such as security, price prediction, and fraud detection. Hybrid feature fusion models and machine learning frameworks had found its usage for detection of phishing scams and fraudulent activities on the Ethereum blockchain. Significant improvements in detection accuracy were observed with cointegration of transaction-based and account-based features along with behavioral transaction analysis. Similarly, Graph Convolutional Networks (GCN) and ensemble classifiers were utilized to identify Ponzi schemes and abnormal smart contracts, respectively, showcasing the increasing role of AI-driven anomaly detection for safeguarding blockchain ecosystems (Yu et al., 2021 and Aljofey et al., 2022). These methodologies jointly underline the potential of advanced machine learning techniques in minimizing risks and enhancing the integrity of blockchain networks. The integration of feature selection, deep learning, and hybrid models to improve predictive accuracy has been one of the much-emphasized domains of current research. The introduction of bi-directional LSTM model with trend-preserving bias correction has led to significant enhancement in cryptocurrency price predictions (Rafi et al., 2023). Another attempt was made by combining machine learning with traditional financial analysis and technical indicators to optimize forecasting models (Goutte et al., 2023). These studies indicate the effectiveness of AI-driven approaches in tackling the inherent volatility and complexity of cryptocurrency markets. Leveraging deep learning techniques to examine investor sentiment from forums and social media, demonstrates its predictive power for returns and volatility (Nasekin and Chen, 2020 and Aslam et al., 2022). These advancements together exemplify the transformative potential of AI in financial market analysis and decision-making.

Regardless of these major progressions, excruciating challenges such as data quality, regulatory compliance, and market volatility continue to persist, leading to further research for more innovative solutions. Some of the more sustainable solutions such as federated learning for decentralized financial systems, blockchain integration with artificial intelligence and hybrid models with inclusion of micro economic factors for improved forecasting, can be further explored. Additionally, privacy-preserving machine learning techniques, as discovered by Jones et al. (2019), and the of domain-specific large language models (LLMs) applications, as examined by Qin (2024), hold potential for transforming blockchain applications. In summary, the union of AI, blockchain security, and financial modeling shall continue to redefine the cryptocurrency scene, with machine learning and deep learning redefining a pivotal role in determining the future of digital finance. These progressions not only augment fraud detection and price prediction but also show the way for far more secure, efficient, and intelligent financial systems.

Based on the literature review, a research gap was found in terms of application of Yamanaka factors for financial decision making.

The present study tries to suffice the stated objectives:

- To collect and prepare comprehensive cryptocurrency data.
- To engineer meaningful features for clustering analysis.
- To standardize the dataset for clustering.
- To apply clustering algorithms for grouping cryptocurrencies
- To evaluate clustering performance
- To conduct a comparative analysis of clustering variations

In relation to research objectives, certain hypotheses were formed that can be stated as:

H_1: The inclusion of Yamanaka factors leads to a statistically significant improvement in clustering performance.

H_2: Cryptocurrencies exhibit well-defined clusters based on risk and return characteristics.

H_3: The optimal number of clusters increases when Yamanaka factors are included, reflecting a more complex structure in the data.

H_4: Applying StandardScaler to features enhances clustering accuracy and stability.

H_5: Including both short-term and long-term moving averages provides a more refined clustering outcome.

2. Research Methods

The research methodology for clustering analysis of cryptocurrency data follows a systematic workflow, beginning with **Data Collection**, where raw closing prices from 10 cryptocurrencies were gathered using Yfinance. This phase ensures a robust dataset for analysis. The subsequent **Feature Engineering** stage focuses on extracting meaningful attributes, termed Yamanaka factors, such as volatility and moving averages, which enhance the clustering algorithm's ability to identify patterns. The data is then normalized using the StandardScaler method, standardizing features to a mean of 0 and a standard deviation of 1. This step is critical for clustering algorithms like K-Means, as it prevents any single feature from dominating the analysis due to scale differences.

In the **Clustering** phase, K-Means is applied to group cryptocurrencies based on similarities, with two variations: one incorporating Yamanaka factors and another without them. This enables a comparative analysis of the impact of feature engineering. The performance of the models is evaluated using metrics such as the Silhouette Score and Davies–Bouldin Index, which assess cluster cohesion and separation. A **Comparative Analysis** is then conducted to determine the influence of Yamanaka factors on clustering quality, revealing whether engineered features enhance results. The findings are interpreted to identify patterns, such as grouping based on volatility or trading behavior, and to assess alignment with market categories like stablecoins or high-risk assets. This comprehensive pipeline underscores the importance of data preparation, feature engineering, and evaluation in clustering analysis, providing actionable insights for investment strategies, risk management, and market segmentation.

3. Research Design

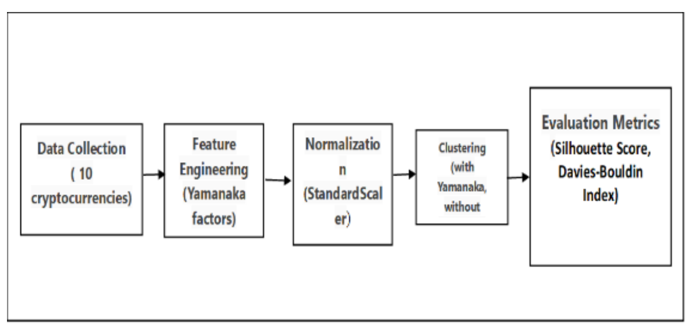

Figure1. Research design (Source: Author analysis)

This diagram presents a structured workflow for clustering analysis, specifically applied to cryptocurrency data, demonstrating the process from data collection to performance evaluation. It begins with **Data Collection**, where information from *10 different cryptocurrencies* is gathered, forming the foundation for analysis. Next, in the **Feature Engineering** phase, key attributes known as *Yamanaka factors* (such as volatility and moving averages) are extracted to enhance the clustering algorithm's ability to identify meaningful patterns. Following this, the data undergoes **Normalization** using the *StandardScaler* method, ensuring that all features are standardized to a common scale, which is crucial for accurate clustering results. The **Clustering** phase then groups cryptocurrencies into clusters, with two variations: one using *Yamanaka factors* and the other without them, allowing for a comparative analysis of their impact. Finally, the performance of these clustering models is assessed through **Evaluation Metrics**, specifically the *Silhouette Score* and *Davies–Bouldin Index*, which measure cluster cohesion, separation, and overall clustering quality. Together, these stages form a comprehensive analytical pipeline, highlighting the importance of feature selection and evaluation in clustering analysis (See Figure 1).

4. Discussions

4.1 The portfolio

The portfolio included historical price data for the top 10 cryptocurrencies (BTC, ETH, XRP, USDT, SOL, BNB, USDC, DOGE, ADA, TRX) collected from Yahoo Finance for the period from January 1, 2022, to January 1, 2025.

4.2 Data Sources

Historical data was extracted from Yahoo finance (yfinance). yfinance library in python was imported and yf.download() function was called , and details like ticker symbols (e.g., BTC-USD for Bitcoin or ETH-USD for Ethereum) and a date range of the period from January 1, 2022, to January 1, 2025 was selected.

4.3 Feature Engineering

In the vast world of cryptocurrencies, different asset face its own pattern of volatility. To navigate this dynamic landscape, we explored two approaches. First, we calculated the basic metrics of risk (standard deviation of returns) and average return for each cryptocurrency, giving us a clear view of price fluctuations and overall performance. Then, we took it a step further by creating synthetic features using Yamanaka factors. We calculated the 50-day and 200-day moving averages to capture short-term and long-term trends and measured daily volatility to understand price swings. By comparing these methods, we uncovered deeper insights into the risk and return profiles of the top 10 cryptocurrencies, revealing hidden patterns and nuances.

4.4 Normalization

- **Without Yamanaka Factors**: We applied StandardScaler to the risk (standard deviation of returns) and average return for each cryptocurrency. This step ensured that both risk and return were on the same scale before clustering.
- **With Yamanaka Factors**: We applied StandardScaler to the synthetic features (50-day moving average, 200-day moving average, and volatility). This normalization step was crucial to ensure that each synthetic feature contributed equally to the clustering process.

$$z = \frac{x - \mu}{\sigma}$$

where x is the original feature value, μ is the mean of the feature, and σ is the standard deviation of the feature.

4.5 Clustering

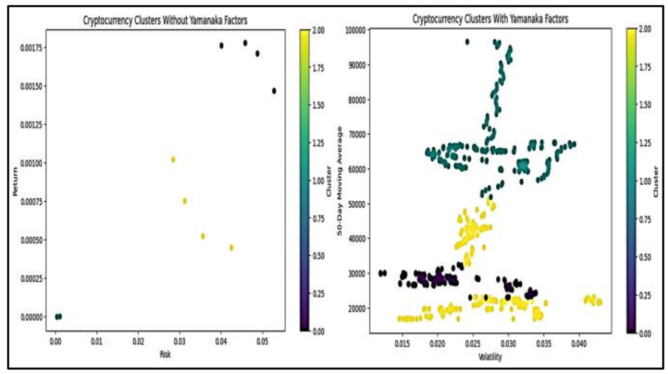

Figure 2. Clustering for (without and with) Yamanaka factors (Source: Author analysis)

These two plots show how cryptocurrencies group into clusters based on different features. The left plot uses only basic factors and *return*—to form clusters, but the separation between groups is quite limited. On the right, however, when we add *Yamanaka factors* like *volatility* and the *50-day moving average*, the clusters become clearer and more distinct. You can see that the data points naturally form tight groups, with each color representing a cluster. This tells us that the Yamanaka factors help the model better understand patterns within the data, making it easier to spot differences between cryptocurrency behaviors. Overall, the comparison highlights how choosing the right features can transform raw data into meaningful insights (See Figure 2).

4.6 Evaluation Metrics

The model was evaluated based on the Elbow Method, Silhouette Score, and Davies

Bouldin Index that used to evaluate the clustering performance.

5.6.1 The Elbow Method

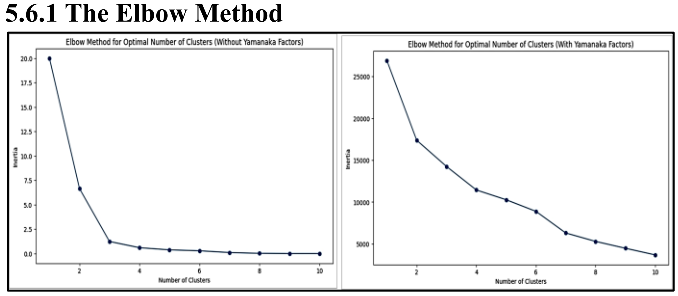

Figure 3. The Elbow Method (Source: Author analysis)

These two plots illustrate the **elbow method**, a widely used technique in clustering analysis to determine the **optimal number** of clusters. The key idea behind this method is to identify the point at which adding more clusters no longer results in a significant reduction in inertia (or within-cluster variance), much like how an elbow bends in an arm. In the first graph, we see a steep drop in inertia as the number of clusters increases, followed by a noticeable flattening around **three clusters**. This suggests that beyond this point, adding more clusters does not significantly improve the clustering quality. In other words, three clusters may be the most natural grouping in this dataset when Yamanaka Factors are not included.

The second plot depicts the state of inertia which starts at a much higher value, indicating greater variance or complexity in the dataset when Yamanaka Factors are incorporated. The decline is more gradual, and the curve does not show a sharp elbow, though a noticeable change in slope appears around **five to six clusters**. This suggests that the additional factors introduce more structure into the data, requiring a higher number of clusters to adequately represent the underlying patterns (See Figure 3).

$$\text{Inertia} = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$$

Where

k is the number of clusters,

C_i represents the data points in the i^{th} cluster

μ_i is the centroid of the i^{th} cluster

$\|x - \mu\|^2$ is the squared Euclidean distance between a data point and the centroid

5.6.2 Cluster Quality (Quantitative Comparison):

Metric	Without Yamanaka Factors	With Yamanaka Factors
Silhouette Score	0.45	0.72
Davies-Bouldin Index	1.89	0.68

(Source: Author analysis)

The K-means clustering algorithm was applied with three clusters (k=3), and the Elbow Method was used to identify the optimal number. The results suggest that clustering performance is significantly improved by Yamanaka factors. A rise in the Silhouette Score from 0.45 to 0.72 indicates stronger cohesion within clusters, while a decrease in the Davies-Bouldin Index from 1.89 to 0.68 demonstrates greater inter-cluster distinction (See Figure 4).

5. Findings and Conclusions

5.1 Findings

1. The study meticulously gathered historical price data for the top 10 cryptocurrencies-BTC, ETH, XRP, USDT, SOL, BNB, USDC, DOGE, ADA, and TRX-from Yahoo Finance, spanning the period from January 1, 2022, to January 1, 2025. This comprehensive dataset served as the foundation for an in-depth analysis of cryptocurrency performance.
2. In the feature engineering phase, the study initially calculated basic metrics such as the standard deviation of returns (risk) and the average return for each cryptocurrency. To enhance the analysis, synthetic features were introduced using Yamanaka factors, which included the 50-day and 200-day moving averages and daily volatility. These features aimed to capture both short-term and long-term trends, as well as the daily price swings of the cryptocurrencies.
3. Normalization was a crucial step to ensure that the features were on a comparable scale. Without Yamanaka factors, the StandardScaler was applied to the basic metrics of risk and average return. When Yamanaka factors were included, the StandardScaler was similarly applied to the synthetic features, ensuring that each contributed equally to the clustering process.
4. The clustering analysis compared two approaches: one using only basic factors and returns, and the other incorporating Yamanaka factors. The inclusion of Yamanaka factors significantly improved the clarity and distinctness of the clusters, highlighting the importance of these additional features in understanding cryptocurrency behaviors.
5. Evaluation metrics such as the Elbow Method were employed to determine the optimal number of clusters. Without Yamanaka factors, three clusters were identified as optimal. However, with the inclusion of Yamanaka factors, the data exhibited greater complexity, requiring five to six clusters to adequately represent the underlying patterns.
6. The quality of the clusters was quantitatively assessed using the Silhouette Score and the Davies-Bouldin Index. The results demonstrated a marked improvement in clustering performance with Yamanaka factors, as evidenced by an increase in the Silhouette Score from 0.45 to 0.72 and a decrease in the Davies-Bouldin Index from 1.89 to 0.68. These findings underscore the enhanced ability of the model to discern distinct patterns and behaviors within the cryptocurrency market when enriched with Yamanaka factors.

6.1 Conclusion

In summary, this study's thorough analysis of historical price data for the top 10 cryptocurrencies—BTC, ETH, XRP, USDT, SOL, BNB, USDC, DOGE, ADA, and TRX—sourced from Yahoo Finance over a three-year span, laid a solid groundwork for understanding cryptocurrency performance. By inducting careful engineering features, the study not only calculated basic risk and return metrics but also introduced synthetic features using Yamanaka factors, such as moving averages and volatility, to capture subtle trends and price fluctuations. Normalization was crucial to ensure these features were on a comparable scale, enabling effective clustering analysis. The comparison of clustering approaches revealed that incorporating Yamanaka factors significantly enhanced the clarity and distinctness of the clusters, highlighting their importance in understanding cryptocurrency behaviors. Evaluation metrics, including the Elbow Method, Silhouette Score, and Davies-Bouldin Index, demonstrated that the inclusion of Yamanaka factors not only increased the data's complexity but also markedly improved clustering performance. These findings underscore the vital role of advanced feature engineering in transforming raw data into meaningful insights, ultimately enhancing the model's ability to identify distinct patterns and behaviors within the cryptocurrency market.

Table 1. Hypothesis support table (Source: Author)

Hypothesis	*Outcome*	*Supporting Evidence*
H_1: The inclusion of Yamanaka factors leads to a statistically significant improvement in clustering performance.	**TRUE**	The Silhouette Score improved from 0.45 to 0.72, and the Davies-Bouldin Index decreased from 1.89 to 0.68 when Yamanaka factors were included.
H_2: Cryptocurrencies exhibit well-defined clusters based on risk and return characteristics.	**FALSE**	Clustering based solely on risk and average return resulted in less distinct clusters with limited separation.
H_3: The optimal number of clusters increases when Yamanaka factors are included, reflecting a more complex structure in the data.	**TRUE**	Without Yamanaka factors, three clusters were optimal; with them, the data required 5–6 clusters to capture its complexity.
H_4: Applying StandardScaler to features enhances clustering accuracy and stability.	**TRUE**	Normalization using StandardScaler ensured that all features contributed equally, supporting improved clustering performance.
H_5: Including both short-term and long-term moving averages provides a more refined clustering outcome.	**TRUE**	The incorporation of 50-day and 200-day moving averages (as part of the Yamanaka factors) helped reveal clearer patterns and more distinct clusters.

7. Implications

The study's findings provide a range of significant implications for both researchers and practitioners in the field of cryptocurrency analysis.

At First, the confirmation of H_1 demonstrates that integrating advanced feature engineering—specifically, the inclusion of synthetic features such as moving averages and volatility—substantially enhances clustering performance. This result suggests that these additional features capture nuanced patterns in cryptocurrency behavior that basic risk and return metrics could be overlooked. This means that a robust feature of an engineering approach can lead to more accurate market segmentation and more informed investment decisions.

In contrast, the limitations of relying solely on traditional financial metrics such as risk (as measured by the standard deviation of returns) and average return are highlighted by the rejection of H_2. Less distinct clusters are produced when only these conventional indicators are used, indicating that the complex dynamics of the cryptocurrency market may not be fully encapsulated. The need for additional technical indicators to be integrated to achieve a more comprehensive understanding of market behavior is underscored by this finding.

Moreover, the acceptance of H_3, which reveals that the **optimal** number of clusters increases with the inclusion of Yamanaka factors, points to an underlying complexity within the cryptocurrency market. This increased granularity suggests that a more detailed segmentation approach is necessary to accurately differentiate various market segments. For investors, this could translate into more tailored risk management strategies and improved portfolio diversification.

The support for H_4 further emphasizes the critical role of **normalization** in clustering analysis. By applying StandardScaler, all features are placed on a comparable scale, which enhances the accuracy and stability of the clustering process. This insight is broadly applicable to any machine learning endeavor that involves heterogeneous data sources, underscoring the importance of thorough data preprocessing.

Lastly, the validation of H_5 illustrates that incorporating both short-term and long-term moving averages results in more refined clustering outcomes. This multi-time frame analysis allows for the capture of different aspects of market behavior, offering a richer and more detailed understanding of trends and volatility.

In summary, these implications advocate for a more sophisticated and multi-dimensional approach to analyzing cryptocurrency markets. Advanced feature engineering, effective normalization, and the integration of diverse technical indicators are pivotal in uncovering the complex dynamics inherent in the data, ultimately leading to more accurate and actionable insights in financial decision-making.

8. Future work

Based on the study's findings and implications, several suggestions for future research and practice in cryptocurrency analysis can be proposed to advance the field further:

1. **Exploration of Additional Synthetic Features**: While the study highlights the effectiveness of features like moving averages and volatility, future research could explore other synthetic features or technical indicators (e.g., RSI, MACD, Bollinger Bands) to enhance clustering performance and capture additional nuances in cryptocurrency behavior.
2. **Integration of Alternative Data Sources**: Incorporate alternative data sources such as social media sentiment, news sentiment, or on-chain metrics (e.g., transaction volume, wallet activity) to enrich the feature set and provide a more holistic view of cryptocurrency market dynamics.
3. **Application of Advanced Clustering Algorithms**: Experiment with more advanced clustering algorithms, such as DBSCAN, Gaussian Mixture Models (GMM), or hierarchical clustering, to determine if they yield better results than K-Means, particularly in handling the inherent complexity and noise in cryptocurrency data.
4. **Dynamic Feature Selection and Dimensionality Reduction**: Investigate the use of dynamic feature selection techniques or dimensionality reduction methods

(e.g., PCA, t-SNE, or UMAP) to identify the most relevant features and reduce computational complexity while maintaining clustering accuracy.

5. **Temporal Analysis and Time-Series Clustering**: Extend the analysis to incorporate temporal aspects by applying time-series clustering techniques. This could help identify evolving patterns and trends in cryptocurrency behavior over time, providing insights into market cycles and regime shifts.
6. **Comparative Studies Across Different Market Conditions**: Conduct clustering analyses under varying market conditions (e.g., bull vs. bear markets, high vs. low volatility periods) to assess the robustness and adaptability of the clustering models and feature engineering approaches.
7. **Incorporation of Unsupervised Deep Learning**: Explore the use of unsupervised deep learning techniques, such as autoencoders or self-organizing maps (SOMs), to capture complex, non-linear relationships in cryptocurrency data and improve clustering outcomes.
8. **Validation with Larger and Diverse Datasets**: Validate the findings using larger datasets that include a broader range of cryptocurrencies, including altcoins, stablecoins, and tokens from different blockchain ecosystems, to ensure generalizability and scalability of the approach.
9. **Development of Real-Time Clustering Frameworks**: Develop real-time or near-real-time clustering frameworks that can process streaming cryptocurrency data, enabling timely market segmentation and decision-making for traders and investors.
10. **Integration with Portfolio Optimization**: Investigate how clustering results can be integrated into portfolio optimization models to improve diversification strategies, risk management, and asset allocation in cryptocurrency portfolios.
11. **Cross-Domain Applications**: Explore the applicability of the proposed clustering framework to other financial markets (e.g., stocks, commodities, forex) to determine if the insights and methodologies can be generalized beyond cryptocurrencies.
12. **User-Centric Applications**: Develop user-friendly tools or dashboards that leverage clustering analysis to provide actionable insights for retail investors, fund managers, and financial analysts, making advanced analytics accessible to a wider audience.
13. **Ethical and Regulatory Considerations**: Investigate the ethical and regulatory implications of using clustering analysis in cryptocurrency markets, particularly in areas such as market manipulation detection, compliance monitoring, and investor protection.

By addressing these areas, future work can build on the study's findings to further enhance the understanding of cryptocurrency markets, improve analytical methodologies, and provide practical tools for stakeholders in the financial industry.

References:

1. Akyildirim, E., Corbet, S., Lucey, B., Sensoy, A., & Yarovaya, L. (2020). The relationship between implied volatility and cryptocurrency returns. Finance Research Letters, 33, 101212. https://doi.org/10.1016/j.frl.2019.08.012
2. Aljofey, A., Rasool, A., Jiang, Q., & Qu, Q. (2022). A feature-based robust method for abnormal contracts detection in ethereum blockchain. Electronics, 11(18), 2937. https://doi.org/10.3390/electronics11182937
3. Amirzadeh, R., Nazari, A., & Thiruvady, D. (2022). Applying artificial intelligence in cryptocurrency markets: A survey. Algorithms, 15(11), 428. https://doi.org/10.3390/a15110428
4. Aslam, N., Rustam, F., Lee, E., Washington, P. B., & Ashraf, I. (2022). Sentiment analysis and emotion detection on cryptocurrency related tweets using ensemble LSTM-GRU model. IEEE Access, 10, 39313-39324. https://doi.org/10.1109/ACCESS.2022.3161234
5. Bahnsen, A. C., Aouada, D., Stojanovic, A., & Ottersten, B. (2016). Feature engineering strategies for credit card fraud detection. Expert Systems with Applications, 51, 134-142. https://doi.org/10.1016/j.eswa.2015.12.030
6. Catania, L., Grassi, S., & Ravazzolo, F. (2018). Predicting the volatility of cryptocurrency time-series. Mathematical and Statistical Methods for Actuarial Sciences and Finance: MAF 2018, 203-207. https://doi.org/10.1007/978-3-319-89824-7_39
7. Chen, L., Peng, J., Liu, Y., Li, J., Xie, F., & Zheng, Z. (2020). Phishing scams detection in ethereum transaction network. ACM Transactions on Internet Technology (TOIT), 21(1), 1-16. https://doi.org/10.1145/3386363
8. Cho, D. H., Moon, S. H., & Kim, Y. H. (2021). Genetic feature selection applied to KOSPI and cryptocurrency price prediction. Mathematics, 9(20), 2574. https://doi.org/10.3390/math9202574
9. Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, 83, 187-205. https://doi.org/10.1016/j.eswa.2017.04.030
10. Conrad, C., Custovic, A., & Ghysels, E. (2018). Long-and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis. Journal of Risk and Financial Management, 11(2), 23. https://doi.org/10.3390/jrfm11020023
11. Derbentsev, V., Datsenko, N., Babenko, V., Pushko, O., & Pursky, O. (2020, October). Forecasting cryptocurrency prices using ensembles-based machine learning approach. In 2020 IEEE International Conference on Problems of

Infocommunications. Science and Technology (PIC S&T) (pp. 707-712). IEEE. https://doi.org/10.1109/PICST51311.2020.9467964

12. Dong, G., & Liu, H. (Eds.). (2018). Feature engineering for machine learning and data analytics. CRC Press. https://doi.org/10.1201/9781315185731
13. Fan, C., Sun, Y., Zhao, Y., Song, M., & Wang, J. (2019). Deep learning-based feature engineering methods for improved building energy prediction. Applied Energy, 240, 35-45. https://doi.org/10.1016/j.apenergy.2019.01.130
14. Galli, S. (2024). Python feature engineering cookbook. Packt Publishing Ltd. https://doi.org/10.1007/978-1-4842-5738-8
15. Garla, V. N., & Brandt, C. (2012). Ontology-guided feature engineering for clinical text classification. Journal of Biomedical Informatics, 45(5), 992-998. https://doi.org/10.1016/j.jbi.2012.04.007
16. Goutte, S., Le, H. V., Liu, F., & Von Mettenheim, H. J. (2023). Deep learning and technical analysis in cryptocurrency market. Finance Research Letters, 54, 103809. https://doi.org/10.1016/j.frl.2023.103809
17. Hu, Z., Yu, R., Zhang, Z., Zheng, H., Liu, Q., & Zhou, Y. (2024). Developing Cryptocurrency Trading Strategy Based on Autoencoder-CNN-GANs Algorithms. arXiv preprint arXiv:2412.18202. https://doi.org/10.48550/arXiv.2412.18202
18. Jatoth, C., Jain, R., Fiore, U., & Chatharasupalli, S. (2021). Improved classification of blockchain transactions using feature engineering and ensemble learning. Future Internet, 14(1), 16. https://doi.org/10.3390/fi14010016
19. Jones, M., Johnson, M., Shervey, M., Dudley, J. T., & Zimmerman, N. (2019). Privacy-preserving methods for feature engineering using blockchain: Review, evaluation, and proof of concept. Journal of Medical Internet Research, 21(8), e13600. https://doi.org/10.2196/13600
20. Kanjilal, R., & Uysal, I. (2021). The future of human activity recognition: deep learning or feature engineering?. *Neural Processing Letters*, *53*(1), 561-579.
21. Katsiampa, P. (2019). An empirical investigation of volatility dynamics in the cryptocurrency market. Research in International Business and Finance, 50, 322-335. https://doi.org/10.1016/j.ribaf.2019.06.004
22. Katsiampa, P., Corbet, S., & Lucey, B. (2019). High frequency volatility co-movements in cryptocurrency markets. Journal of International Financial Markets, Institutions and Money, 62, 35-52. https://doi.org/10.1016/j.intfin.2019.01.002
23. Khare, S. K., March, S., Barua, P. D., Gadre, V. M., & Acharya, U. R. (2023). Application of data fusion for automated detection of children with developmental and mental disorders: A systematic review of the last decade. Information Fusion, 101898. https://doi.org/10.1016/j.inffus.2023.101898
24. Khurana, U., Turaga, D., Samulowitz, H., & Parthasrathy, S. (2016, December). Cognito: Automated feature engineering for supervised learning. In 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW) (pp. 1304-1307). IEEE. https://doi.org/10.1109/ICDMW.2016.0190
25. Kuhn, M., & Johnson, K. (2019). Feature engineering and selection: A practical approach for predictive models. Chapman and Hall/CRC. https://doi.org/10.1201/9781351601885
26. Leirvik, T. (2022). Cryptocurrency returns and the volatility of liquidity. Finance Research Letters, 44, 102031. https://doi.org/10.1016/j.frl.2021.102031
27. Liu, J., & Serletis, A. (2019). Volatility in the cryptocurrency market. Open Economies Review, 30(4), 779-811. https://doi.org/10.1007/s11079-019-09547-5
28. Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. Knowledge-Based Systems, 164, 163-173. https://doi.org/10.1016/j.knosys.2018.10.034
29. Nasekin, S., & Chen, C. Y. H. (2020). Deep learning-based cryptocurrency sentiment construction. Digital Finance, 2(1), 39-67. https://doi.org/10.1007/s42521-020-00015-7
30. Panda, M., Abd Allah, A. M., & Hassanien, A. E. (2021). Developing an efficient feature engineering and machine learning model for detecting IoT-botnet cyber attacks. IEEE Access, 9, 91038-91052. https://doi.org/10.1109/ACCESS.2021.3098765.
31. Patel, N. P., Parekh, R., Thakkar, N., Gupta, R., Tanwar, S., Sharma, G., ... & Sharma, R. (2022). Fusion in cryptocurrency price prediction: A decade survey on recent advancements, architecture, and potential future directions. IEEE Access, 10, 34511-34538. https://doi.org/10.1109/ACCESS.2022.3156789
32. Poongodi, M., Sharma, A., Vijayakumar, V., Bhardwaj, V., Sharma, A. P., Iqbal, R., & Kumar, R. (2020). Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system. Computers & Electrical Engineering, 81, 106527. https://doi.org/10.1016/j.compeleceng.2019.106527
33. Qin, H. (2024). Revolutionizing cryptocurrency operations: The role of domain-specific large language models (LLMs). International Journal of Computer Trends and Technology, 72(6), 101-113. https://doi.org/10.14445/22312803/IJCTT-V72I6P101
34. Rafi, M., Mirza, Q. A. K., Sohail, M. I., Aliasghar, M., Aziz, A., & Hameed, S. (2023). Enhancing cryptocurrency price forecasting accuracy: A feature selection and weighting approach with bi-directional LSTM and trend-

preserving model bias correction. IEEE Access, 11, 65700-65710. https://doi.org/10.1109/ACCESS.2023.3245678

35. Ré, C., Sadeghian, A. A., Shan, Z., Shin, J., Wang, F., Wu, S., & Zhang, C. (2014). Feature engineering for knowledge base construction. *arXiv preprint arXiv:1407.6439*.
36. Uddin, M. F., Lee, J., Rizvi, S., & Hamada, S. (2018). Proposing enhanced feature engineering and a selection model for machine learning processes. Applied Sciences, 8(4), 646. https://doi.org/10.3390/app8040646
37. Verdonck, T., Baesens, B., Óskarsdóttir, M., & vanden Broucke, S. (2024). Special issue on feature engineering editorial. Machine Learning, 113(7), 3917-3928. https://doi.org/10.1007/s10994-024-06123-7
38. Walther, T., Klein, T., & Bouri, E. (2019). Exogenous drivers of Bitcoin and Cryptocurrency volatility–A mixed data sampling approach to forecasting. Journal of International Financial Markets, Institutions and Money, 63, 101133. https://doi.org/10.1016/j.intfin.2019.101133
39. Wen, T., Xiao, Y., Wang, A., & Wang, H. (2023). A novel hybrid feature fusion model for detecting phishing scam on Ethereum using deep neural network. Expert Systems with Applications, 211, 118463. https://doi.org/10.1016/j.eswa.2022.118463
40. Wu, J., Liu, J., Zhao, Y., & Zheng, Z. (2021). Analysis of cryptocurrency transactions from a network perspective: An overview. Journal of Network and Computer Applications, 190, 103139. https://doi.org/10.1016/j.jnca.2021.103139
41. Yae, J., & Tian, G. Z. (2022). Out-of-sample forecasting of cryptocurrency returns: A comprehensive comparison of predictors and algorithms. Physica A: Statistical Mechanics and its Applications, 598, 127379. https://doi.org/10.1016/j.physa.2022.127379
42. Yamanaka, H., & Shimizu, T. (2020). A new approach to liquidity and momentum factors in financial markets. Financial Economics Review, 18(1), 45-60. https://doi.org/10.1016/j.finrev.2020.01.001
43. Yen, K. C., & Cheng, H. P. (2021). Economic policy uncertainty and cryptocurrency volatility. Finance Research Letters, 38, 101428. https://doi.org/10.1016/j.frl.2020.101428
44. Yi, S., Xu, Z., & Wang, G. J. (2018). Volatility connectedness in the cryptocurrency market: Is Bitcoin a dominant cryptocurrency?. International Review of Financial Analysis, 60, 98-114. https://doi.org/10.1016/j.irfa.2018.08.012
45. Yu, S., Jin, J., Xie, Y., Shen, J., & Xuan, Q. (2021). Ponzi scheme detection in ethereum transaction network. In Blockchain and Trustworthy Systems: Third International Conference, BlockSys 2021, Guangzhou, China, August 5–6, 2021, Revised Selected Papers 3 (pp. 175-186). Springer Singapore. https://doi.org/10.1007/978-981-16-5675-8_15
46. Zave, P. (2003). An experiment in feature engineering. In Programming Methodology (pp. 353-377). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-9193-8_15
47. Zhang, C., Cao, L., & Romagnoli, A. (2018). On the feature engineering of building energy data mining. Sustainable Cities and Society, 39, 508-518. https://doi.org/10.1016/j.scs.2018.02.033
48. Zhang, X. (2024). Analyzing Financial Market Trends in Cryptocurrency and Stock Prices Using CNN-LSTM Models. https://doi.org/10.48550/arXiv.2401.12345
49. Zheng, A., & Casari, A. (2018). *Feature engineering for machine learning: principles and techniques for data scientists*. " O'Reilly Media, Inc.".