2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Portfolio Diversification with Clustering Techniques

Joy Dip Das, Sulalitha Bowala, Ruppa K. Thulasiram, Aerambamoorthy Thavaneswaran
Department of Computer Science
University of Manitoba
Winnipeg, Canada
dasj@myumanitoba.ca;
bowalams@myumanitoba.ca (Department of Statistics);
tulsi.thulasiram@umanitoba.ca;
Aerambamoorthy.Thavaneswaran@umanitoba.ca (Department of Statistics)

***Abstract*—Diversifying asset allocation is a crucial aspect of building a profitable portfolio. The resiliency of the portfolio depends on the optimization techniques as well as algorithms used in the asset allocation. Clustering techniques would help in designing a diversified portfolio. This study investigates the resiliency of different traditional and recently proposed data-driven portfolio techniques in conjunction with four clustering techniques under varying market conditions. The novelty of the study is to present a resilient portfolio optimization using DBSCAN and Affinity Propagation clustering techniques.**

***Index Terms*—Portfolio Diversification, Resilience, Clustering, DBSCAN, Affinity Propagation**

I. Introduction

The global financial crisis of 2008 posed a challenge to the risk management abilities of market participants and raised concerns about portfolio risk management methods and practices [1]. Resiliency in portfolio construction is very important in averting such drastic economic collapse. The Mean-Variance Optimization (MVO) model [2] is widely accepted as the standard for assessing portfolio performance. MVO enables the creation of an efficient portfolio that strives to achieve the highest return for a unit of non-systematic risk or the lowest non-systematic risk for a unit of return. Linear programming is used to find the optimal weights in Sharpe ratio Optimization(SRO). Perold [3] focused on optimizing big-scale portfolios and implemented a fundamental structured model that incorporated variables such as transaction costs and the covariance matrix to achieve a reasonably accurate estimation. Wang and Aste [4] have implemented the Inverse Covariance Clustering (ICC) technique to identify market states. They have incorporated these states into a dynamic approach to tackle the challenge posed by the constant fluctuations in the market.

ML algorithms are extensively investigated in this domain. The study of ML and DL algorithms, in conjunction with the Markowitz portfolio model, aims to determine optimal portfolio weights for efficient portfolio construction. Awoye [5] investigated the use of machine learning techniques to enhance portfolio returns by applying the minimum variance model of Markowitz. Chen et al. [6] employed the XGBoost algorithm and Firefly algorithm in conjunction with the MVO model to forecast and identify high-performing stocks for the purpose of resilient portfolio optimization. Du [7] has employed a comparable methodology, utilizing ML algorithms like Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) for stock price prediction, followed by the application of MVO to construct portfolios. There is an ongoing exploration in portfolio construction regarding the utilization of deep learning (DL) algorithms, including Autoencoder and LSTM models [8] [9]. Nevertheless, a significant limitation lies in the fact that the majority of asset returns do not adhere to normal distributions. Hence, the utilization of sample percentiles is for studying portfolio optimization using Value-at-Risk (VaR) as a risk metric [10]. Addressing this constraint involves the utilization of a data-driven algorithm for generating random weights (RWA) and employing a portfolio return distribution based on sign correlations in the context of fuzzy portfolio optimization.

Diversifying asset allocation is a critical aspect of constructing a robust portfolio. Choueifaty and Coignard [11] introduced the concept of the 'diversification ratio' to quantify the level of diversity in a portfolio. Ibanez [12] proposed a portfolio construction technique that emphasizes diversification. This approach employs singular value decomposition to identify underlying factors and utilizes hierarchical agglomerative clustering to address implementation challenges associated with the method. Clustering algorithms can be used in diversifying financial assets and constructing risk-averse portfolios. Li [13] applied K-means clustering to enhance the performance of the global minimum variance (GMV) portfolio. The approach also incorporates regularization within the autoregressive shrinkage model. Aslam et al. [14] have incorporated a three-factor analysis to cluster stocks and identify clusters comprising the best-performing stocks for portfolio construction. Pranata et al. [15] employed clustering techniques such as the k-means and DBSCAN clustering algorithms to cluster the IDX (Indonesia Stock Exchange) companies. These clustering techniques are commonly employed in stock forecasting and recommendation systems [16] as well as to optimize portfolio construction and risk management [17].

Thavaneswaran et al. [18] introduced a data-driven approach for portfolio risk assessment to overcome the limitations pf traditional techniques. Building upon prior work, Bowala and Singh [19] introduced a data-centric method for composing a

978-0-7381-4408-5/23/$31.00

portfolio comprising conventional stocks and cryptocurrencies (cryptos). They explored statistical risk measures and put forward a range of statistical correlations for assessing portfolio composition utilizing these metrics.

The current study investigates the effectiveness of four distinct clustering algorithms in facilitating diversification within portfolio construction. The study encompasses both traditional portfolio construction techniques, including MVO, CPO, and SRO, as well as the Data-Driven approach or portfolio construction. The objective of this research is to enhance portfolio optimization methodologies through the utilization of diversification via clustering techniques. This study also aims to enhance the performance of data-driven portfolio construction by incorporating diversification through the utilization of clustering techniques. Additionally, it investigates less explored clustering algorithms, namely Density-based spatial clustering of applications with noise (DBSCAN) and affinity propagation, in the context of portfolio optimization. By applying clustering algorithms to enhance portfolio diversification, valuable insights are obtained regarding the performance of different algorithms under varying market conditions. The study incorporates stock market data from multiple sectors, such as technology, consumer goods, banking, energy, automobiles, and pharmaceuticals. The findings of this study primarily focused on assessing the efficacy of diverse clustering algorithms in identifying profitable portfolio returns amidst volatile market conditions.

In Section II, the document outlines clustering algorithm methodologies and portfolio construction approaches, encompassing both traditional and data-driven methods. Section III encompasses the detailed experimental procedures and outcomes, while Section IV offers closing remarks and conclusions.

II. Methodologies

Conventional portfolio optimization techniques such as Mean-Variance Optimization (MVO) and Sharpe Ratio Optimization (SRO) aim to strike a balance between risk and return by solving specific objective functions. However, these approaches often overlook the importance of diversification in constructing risk-averse portfolios. Consequently, they underperform in volatile market conditions. This study investigates the impact of diversification implemented through clustering techniques across different market conditions. By applying various clustering methods to raw financial asset data, portfolios comprising assets from diverse clusters are constructed. These portfolios are then evaluated using both traditional and data-driven portfolio optimization techniques under varying market conditions. The distinct functionalities of different clustering techniques necessitate the identification of optimal combinations with portfolio optimization methods to propose a resilient diversification approach.

A. Clustering Algorithms for diversification

This study employs four different clustering techniques to facilitate stock diversification within portfolios. Specifically, the less extensively explored methods of DBSCAN and affinity propagation are studied in this work. The objective is to construct portfolios by selecting unique stocks from each cluster, thereby enhancing diversification strategies.

a) K-means clustering algorithm: It is a prominent unsupervised ML approach utilized for partitioning a given dataset into K distinct clusters. This technique aims to minimize the collective sum of squared distances between individual data points and their corresponding cluster centroids. The Within Cluster Sum of Squares (WCSS) method is commonly employed to determine the suitable number of clusters.

b) Hierarchical Clustering Algorithm: It is an unsupervised ML algorithm that groups data points into hierarchical clusters, considering their similarity or dissimilarity. This method establishes a hierarchical arrangement of clusters, often represented as a dendrogram, by iteratively merging or splitting clusters. In our study, we leveraged the elbow plot technique, obtained by converting the dendrogram into a dissimilarity matrix. By utilizing the WCSS criterion, we ascertained the ideal cluster quantity essential for our clustering investigation.

c) DBSCAN: It is an unsupervised ML algorithm that identifies clusters based on the concept of density. DBSCAN effectively employs two crucial parameters: epsilon (ϵ), which governs the maximum allowable distance between neighboring points, and minPts, which specifies the minimum number of points required to establish a dense region. The algorithm also adeptly detects noise or outlier points, which lie in regions of diminished density and fail to satisfy the density criteria. In our study, we employed the Reachability distance metric to determine the optimal value of epsilon for the DBSCAN clustering algorithm.

d) Affinity Propagation: Affinity Propagation is an influential clustering algorithm that operates in an unsupervised manner that effectively identifies exemplars within a given dataset, allowing the data points to "vote" on each other to determine their representative exemplar. The algorithm leverages the notion of message passing, wherein messages are exchanged between data points to ascertain the most suitable exemplar for each point. By iteratively updating these messages, Affinity Propagation converges to a set of exemplars that effectively represent the dataset. The robustness of the algorithm lies in its capacity to autonomously ascertain the cluster quantity without necessitating prior knowledge. The versatility of this algorithm allows it to handle datasets with varying cluster sizes. This algorithm has been previously used in several applications [20] [21].

B. Traditional Portfolio Construction

Traditional portfolio optimization techniques focus on maximizing profits while minimizing risk or maximizing the Sharpe ratio but overlook important statistical characteristics such as the presence of non-normal market behavior.

a) Equal Weighted Portfolio Optimization (E.W) and Inverse Volatility Weighted Portfolio Optimization (I.V.W): In a portfolio utilizing the equal-weighted portfolio optimization technique, every stock is accorded uniform treatment, and an identical proportion of weight is allocated to each individual stock. Specifically, the weight of each stock is calculated as $W_i = \frac{1}{N}$.

The inverse volatility-weighted portfolio assigns weights to assets based on the reciprocal of their volatility. Specifically, the weight of each asset is calculated as $W_i = \frac{1}{\sigma} \cdot \frac{1}{\sum_j \frac{1}{\sigma_j}}$. Within this mathematical expression, σ_i denotes the volatility associated with the i-th asset. We employ methodologies encompassing equal-weighting and inverse volatility-weighting to compute the portfolio weights. Subsequently, the yearly mean returns, annual risk, and diversification ratio are computed.

The return from a portfolio is specified as follows: Let P consist of assets $A_1, A_2, \ldots, A_N$ with corresponding weights $W_1, W_2, \ldots, W_N$ and $R_1, R_2, \ldots, R_N$ representing the asset returns. The return from portfolio, r is determined by the summation of weighted individual asset returns, expressed as $R_p = W_1R_1 + W_2R_2 + \ldots + W_NR_N = \sum_{i=1}^{N} W_iR_i$. For T_y trading days (e.g., 251 days), the annualized return of the portfolio can be computed as $R_{Ann} = \left(\sum_{i=1}^{N} W_iR_i\right) T_y$. The portfolio's risk, denoted by σ, is calculated as $\sigma = \sqrt{\sum_i \sum_j W_iW_j\sigma_{ij}}$.

The variance-covariance matrix of returns denoted as σ_{ij} between assets i and j is a crucial element. The annualized portfolio risk in percentage, σ_{Ann}, can be calculated as: $\sigma_{Ann} = \sqrt{\left(\sum_i \sum_j W_iW_j\sigma_{ij}\right) T_y}$. Moreover, the diversification index is computed using the following formula: $(\bar{\sigma}\bar{W}')/(\sqrt{\bar{W}\Sigma\bar{W}'})$. In this equation, $\bar{\sigma}$ refers to the vector of standard deviation, $\sqrt{\bar{W}}$ represents the weights vector, and Σ signifies the variance-covariance matrix of returns.

b) Mean-Variance Optimization (MVO): The primary goals of MVO are to boost the anticipated return of a portfolio and reduce its overall risk. This model is expressed as follows: $\min - \sum_{i=1}^{N} W_iR_i + \sqrt{\sum_i \sum_j W_iW_j\sigma_{ij}}$ given that $\sum_{i=1}^{N} W_i = 1, , 0 \leq W_i \leq 1$ where W_i is portfolio weight, R_i is the expected return of portfolio, σ_{ij} is the returns' variance covariance, and N is the total number of assets.

c) Constrained Portfolio Optimization (CPO): To construct an optimally balanced portfolio, an investor with moderate risk aversion customizes the Mean-Variance Optimization model to create the Constrained Portfolio Optimization model, considering specific preferences and requirements. In this research, the CPO model is employed to calculate ideal weights: 40% in stocks with high volatility, 60% in stocks with low volatility, and 1% of total capital in both such categories. Some of the assets might remain uninvested, and the allocation of stocks with high volatility is capped at 10%. The sum of the weights is one, ensuring full capital utilization. The mathematical representation of CPO is: $\min - \sum_{i=1}^{N} W_iR_i + \sum_i \sum_j W_iW_j\sigma_{ij}$ when $\sum_{i=1}^{N} W_i = 1$, $0.01 \leq \sum_{i \in \text{LowVolatility}} W_i \leq 0.6$, $0.01 \leq \sum_{j \in \text{HighVolatility}} W_j \leq 0.4$, $0 \leq W_i \leq 1$ (for $i \in$ LowVolatility), $0 \leq W_j \leq 0.1$ (for $j \in$ HighVolatility).

d) Sharpe Ratio Optimization (SRO): The mathematical representation for SRO is:
$\max \left(\frac{\sum_{i=1}^{N} W_iR_i - R_f}{\sqrt{\sum_i \sum_j W_iW_j\sigma_{ij}}} \right)$, subject to $\sum_{i=1}^{N} W_i = 1$ and $0 \leq W_i \leq 1$.

The investment strategy's objective function contains two parts: the numerator signifies the excess returns of the investment in comparison to an asset that is risk-free, denoted as R_f, and the denominator denotes the risk in the investment. The main goal is to optimize the Sharpe ratio.

C. Data Driven Portfolio Construction

In [18], two novel estimation functions are introduced, which have smaller variances and are utilized to establish a new portfolio risk measure that incorporates higher-order moments. The corresponding volatility correlation ($\rho_{P,\text{vol}}$) and sign correlation ($\rho_{P,\text{sgn}}$) are defined as follows:

$\rho_{P,sgn} = \text{Corr}(Sgn(R_P - \mu_p), R_P - \mu_p)$ and $\rho_{P,vol} = \text{Corr}(|R_P - \mu_p|, (R_P - \mu_p)^2)$ where μ_p is portfolio expected return.

The portfolio's expected return is represented by μ_p, while skewness is denoted by $\tilde{\mu_3}$, and κ represents kurtosis. F(x) is the inverse of the cumulative distribution function (CDF) of portfolio return. $\tilde{R}_p$ and $\hat{\sigma}_P$ are estimates of the portfolio's expected return and standard deviation (SD), respectively, based on the previous l observation. The portfolio's mean absolute deviation (MAD) estimate is calculated using the following formula:
$\text{MAD}_P = 2\hat{\rho}P, \text{sign}\hat{\sigma}_P \sqrt{F(\bar{R}_P)(1 - F(\bar{R}_P))}$.
The calculation of the portfolio volatility estimate using volatility correlation reduction (VEV) is determined by the following formula: $\text{VEV}_P = \sqrt{1 - \hat{\rho}^2_{P,\text{vol}}}\hat{\sigma}_P$

The computation of the portfolio volatility estimate using sign correlation reduction (VES) is determined by the following formula: $\text{VES}_P = \sqrt{1 - \hat{\rho}^2_{P,\text{sign}}}\hat{\sigma}_P$. The calculation of the portfolio volatility estimate using both sign and volatility correlation reduction (VESV) is determined by the following formula- $\text{VESV}_P = \sqrt{(1 - \hat{\rho}^2_{P,sign})(1 - \hat{\rho}^2_{P,vol})}\hat{\sigma}_P$.

III. Experiments and Results

A. Exploratory Analysis

Our findings indicate that the year 2020 exhibited substantial profitability, contrasting with the year 2022 wherein a significant proportion of stocks yielded negative mean returns. For example TESLA (TSLA), throughout the year 2020 showed an upward trend but in 2022, it dropped continuously.

Table I summarizes the cluster number for each year using different clustering algorithms. Different clustering algorithms function differently resulting in varying cluster numbers for different years. For each year we have considered two different portfolios. From each cluster, we have selected the stock with the highest mean return to be included in the **K-01** portfolio,

	kmeans	hierarchical	DBSCAN	Affinity Propagation
2019	12	12	13	9
2020	12	10	14	6
2021	12	12	15	8
2022	12	8	15	8

TABLE I
NUMBER OF CLUSTERS USING DIFFERENT CLUSTERING TECHNIQUES

whereas the stock with the second highest mean return is included in the **K-02** portfolio.

B. Clustering on Equal Weighted and Inverse Volatility Weighted portfolio construction

Although diversification with clustering algorithms improved the performance of simple E.W and I.V.W, they still failed in risky years like 2022. The affinity propagation clustering algorithm produced the best annual profits in 2020, but all clustering algorithms failed to generate positive returns in 2022. Table II presents the performance of different clustering algorithms in conjunction with E.W and I.V.W portfolio construction techniques in different years. E.W and I.V.W are not reliable in risky years like 2022.

Year	Clustering Methods	Equal Weighted				Inverse Volatility Weighted			
		K-01		K-02		K-01		K-02	
		Annual Risk	Annual Return	Annual Risk	Annual Return	Annual Risk	Annual Return	Annual Risk	Annual Return
2019	kmeans	15.91%	25.90%	15.49%	23.81%	14.19%	27.14%	13.37%	23.11%
2020		38.74%	50.37%	37.75%	33.89%	36.59%	44.04%	34.84%	26.76%
2021		17.27%	48.76%	16.24%	42.48%	15.07%	43.54%	13.50%	37.12%
2022		28.87%	-18.75%	29.60%	-23.33%	25.02%	-6.90%	26.59%	-14.82%
2019	Hierarchical	15.91%	25.90%	15.36%	23.31%	14.19%	27.14%	12.96%	22.65%
2020		38.95%	63.19%	37.78%	43.59%	36.52%	55.96%	34.47%	34.80%
2021		16.64%	48.50%	15.76%	42.01%	13.90%	42.02%	13.14%	36.09%
2022		28.55%	-18.02%	29.73%	-24.10%	23.29%	-2.47%	25.67%	-12.72%
2019	DBSCAN	16.56%	27.80%	16.73%	25.70%	15.25%	29.08%	15.29%	26.17%
2020		38.56%	39.53%	38.81%	37.13%	34.75%	35.52%	34.69%	32.62%
2021		16.49%	42.30%	15.95%	40.02%	15.14%	38.76%	14.26%	36.41%
2022		27.37%	-16.38%	21.95%	-69.18%	24.11%	-6.13%	15.02%	-22.80%
2019	Affinity Propagation	17.68%	24.74%	16.99%	20.88%	15.35%	25.83%	14.68%	21.59%
2020		42.83%	85.54%	42.25%	68.44%	39.55%	77.08%	40.09%	56.36%
2021		20.54%	59.04%	20.65%	51.37%	18.46%	56.34%	16.96%	44.75%
2022		28.55%	-18.02%	29.73%	-24.10%	23.29%	-2.47%	25.67%	-12.72%

TABLE II
PERFORMANCE USING DIFFERENT CLUSTERING ALGORITHMS FOR E.W AND I.V.W CONSTRUCTIONS

This study delves into the application of Mean-Variance Optimization (MVO) and Sharpe Ratio Optimization (SRO) in portfolio management. Specifically, the influence of clustering techniques on these strategies are investigated, analyzing their efficacy in enhancing diversification and identifying profitable opportunities (see subsections C and D)

C. Clustering with MVO and CPO

Figures 1 and 2 showcase MVO's effectiveness with diverse clustering algorithms for 2020 and 2022, revealing consistent portfolio outcomes on the efficient frontier. For 2022, the clustering algorithms demonstrate comparable performance characteristics.

Diverse clustering techniques notably boosted conventional MVO performance, resulting in significant 2020 profitability. Impressively, this diversification sustained profits even during the high-risk year of 2022. For 2022, Figure 2 highlights efficient frontiers using k-means clustering. Other algorithms also display similar frontiers.

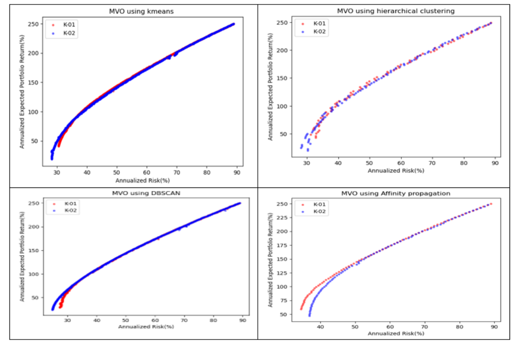

Fig. 1. MVO with different clustering algorithms for 2020

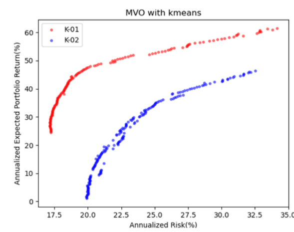

Fig. 2. MVO with kmeans clustering algorithm for 2022

1) Constrained Portfolio Optimization (CPO): CPO extends MVO by accommodating risk-averse investors' constraints, shaping a customized framework. Clustering algorithms showed similar performance in MVO, but CPO indicated less favorability for affinity propagation, especially during 2020 and 2022. Observations from Figure 3 highlight suboptimal performance of CPO with affinity propagation in 2020 compared to kmeans, suggesting limited effectiveness of affinity propagation for CPO-based portfolios.

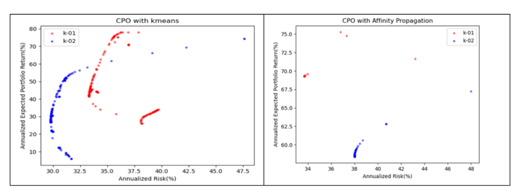

Fig. 3. CPO with kmeans clustering algorithm(left) and Affinity Propagation(right) in 2020

The study revealed that in the tumultuous market of 2022, portfolios created using DBSCAN and k-means clustering algorithms outperformed those using hierarchical and affinity propagation algorithms (Figure 4).

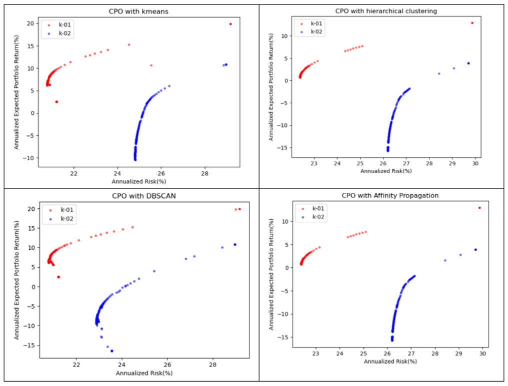

Fig. 4. CPO with clustering algorithms for 2022

D. Clustering with SRO

Sharpe Ratio Optimization maximizes risk-adjusted return using excess return to volatility ratio as a key metric. The application of diverse clustering techniques significantly improved Sharpe Ratio Optimization performance. Notably, for 2020, k-means demonstrated superior risk-adjusted profitability compared to other clustering algorithms, as observed in Table III.

Year	Clustering Methods	SHARPE RATIO OPTIMIZATION					
		K-01			K-02		
		Annual Risk	Annual Return	Sharpe Ratio	Annual Risk	Annual Return	Sharpe Ratio
2019	kmeans	13.77%	42.63%	2.88	14.41%	42.51%	2.74
2020		58.88%	171.31%	2.86	14.41%	42.51%	2.74%
2021		15.16%	57.62%	3.62	13.36%	51.93%	3.66
2022		19.29%	46.57%	2.26	26.58%	38,65%	1.34
2019	Hierarchical	13.77%	42.63%	2.88	13.26%	40.06%	2.80%
2020		58.81%	171.08%	2.83	59.99%	172.80%	2.83
2021		14.42%	55.06%	3.61	14.25%	54.97%	3.65
2022		19.45%	46.92%	2.26	25.54%	36.87%	1.33
2019	DBSCAN	14.68%	45.02%	2.86	14.91%	43.96%	2.75
2020		57.40%	166.76%	2.85	55.57%	161.31%	2.85
2021		16.07%	60.67%	3.59	14.21%	54.04%	3.65
2022		19.05%	45.99%	2.26	25.01%	36.35%	1.33
2019	Affinity Propagation	14.65%	44.08%	2.8	14.81%	36.87%	2.29
2020		62.33%	180.88%	2.85	59.61%	172.66%	2.85
2021		16.89%	63.45%	3.58	17.65%	55.12%	2.95
2022		19.45%	46.93%	2.26	25.54%	36.87%	1.33

TABLE III
SRO WITH DIFFERENT CLUSTERING TECHNIQUES

E. Clustering with Data-driven(DD) approach

In traditional portfolio strategies, non-normality in financial data is overlooked. Data-driven approaches consider this, optimizing portfolios with data-centric risk metrics. This study investigates enhancing these techniques with diversification and recommends effective clustering methods for robust portfolio optimization, particularly in volatile markets.

1) DD without Clustering: For 2022, the data-driven approach without clustering algorithms showed limited profitability, as illustrated in Figure 5 but integrating clustering algorithms not only bolstered performance that year but also improved overall performance across other years. This study highlights varying effectiveness among clustering algorithms in enhancing portfolio construction techniques, particularly in risk-prone years like 2022, emphasizing a data-driven assessment.

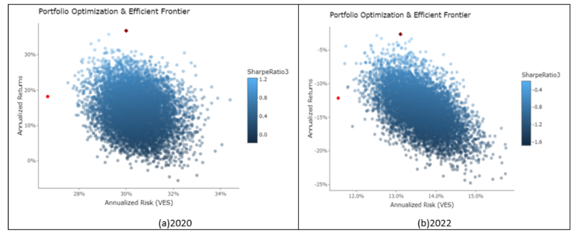

Fig. 5. Data-driven portfolio optimization without clustering

2) DD with Clustering: The integration of diversification through various clustering techniques significantly enhanced the profitability of the data-driven portfolio construction technique.

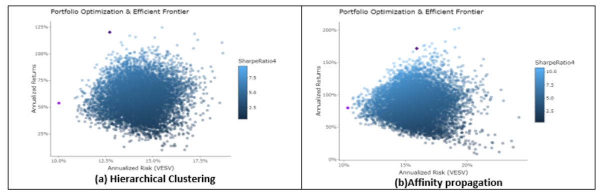

Fig. 6. Data-driven approach with Hierarchical (left) and Affinity Propagation(right) clustering algorithms for 2020

For 2020, the utilization of affinity propagation and hierarchical clustering showcased enhanced performance in diversification compared to alternative algorithms, as supported by Figure 6, with affinity propagation particularly excelling in constructing profitable portfolios during that year.

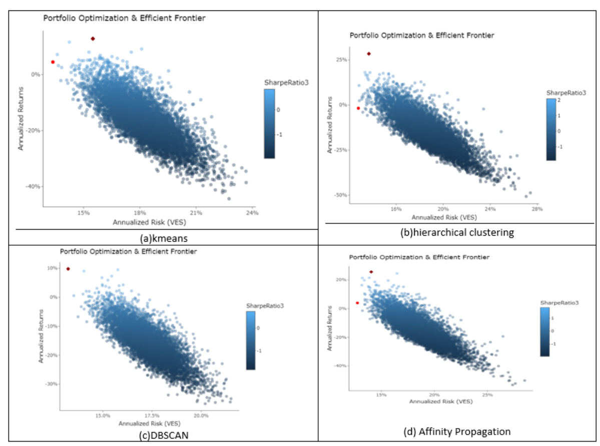

Fig. 7. Data-driven approach with different clustering algorithms for 2022

Figure 7 illustrates the superior performance of hierarchical clustering and affinity propagation-based portfolio construction for 2022 within a data-driven approach, highlighting the potential of affinity propagation for enhanced diversification and portfolio robustness.

IV. Discussions and Conclusions

This study underscores the pivotal role of diversification through clustering techniques in improving portfolio resilience. Analyzing various clustering methods and portfolio construction approaches reveals that portfolio performance is contingent on the choice of clustering technique and prevailing market conditions. Optimal diversification, crucial for portfolio resilience, necessitates a careful selection of assets from disparate clusters. Different clustering methods exhibit distinct strengths, with affinity propagation clustering synergizing effectively with traditional approaches like equal-weighted (E.W) and inverse volatility-weighted (I.V.W) optimization during profitable periods. Mean-Variance Optimization (MVO) demonstrates consistent effectiveness across clustering techniques, while k-means clustering stands out in Sharpe Ratio Optimization (SRO) during favorable market conditions. The integration of data-driven (DD) portfolio construction techniques, notably in conjunction with affinity propagation clustering, remarkably enhances portfolio performance. These insights hold pragmatic implications for both investors and portfolio managers, emphasizing the advantages of diversification facilitated by clustering techniques in constructing resilient portfolios. Future research should further investigate alternative clustering methods and incorporate additional risk metrics to provide a comprehensive understanding of resilient portfolio construction.

ACKNOWLEDGEMENT

The first author acknowledges financial support from Professor Thulasiram and the first two authors acknowledge the University of Manitoba Graduate Fellowship (UMGF) and Graduate Enhancement of Tri-agency Stipends (GETS), University of Manitoba. The last two authors acknowledge the Discovery Grants from the Natural Sciences and Engineering Research Council (NSERC) Canada.

REFERENCES

[1] Appel, D., and Grabinski, M. “The origin of financial crisis: A wrong definition of value.” Portuguese Journal of Quantitative Methods, 2, 33-51, 2011

[2] Ben Salah, H., De Gooijer, J. G., Gannoun, A., and Ribatet, M. (2018). Mean–variance and mean–semivariance portfolio selection: a multivariate nonparametric approach. Financial Markets and Portfolio Management, 32(4), 419-436. https://doi.org/10.1007/s11408-018-0317-4

[3] Tola, V., Lillo, F., Gallegati, M., and Mantegna, R. N. “Cluster analysis for portfolio optimization.” Journal of Economic Dynamics and Control, 32(1), 235-258, 2008.

[4] Wang, Y., and Aste, T. (2023). Dynamic portfolio optimization with inverse covariance clustering. Expert Systems with Applications, 213(Part A), 118739. https://doi.org/10.1016/j.eswa.2022.118739

[5] Awoye, O. A. “Markowitz minimum variance portfolio optimization using new machine learning methods” (Doctoral dissertation, (UCL) University College London), 2016.

[6] Chen, W., Zhang, H., Mehlawat, M. K., and Jia, L. (2021). Mean-variance portfolio optimization using machine learning-based stock price prediction. Applied Soft Computing, 100, 106943. https://doi.org/10.1016/j.asoc.2020.106943

[7] Du, J. (2022). Mean-variance portfolio optimization with deep learning-based forecasts for cointegrated stocks. Expert Systems with Applications, 201, 117005. https://doi.org/10.1016/j.eswa.2022.117005

[8] Ma, Y., Wang, W., and Ma, Q. (2023). A novel prediction based portfolio optimization model using deep learning. Computers and Industrial Engineering, 177, 109023. https://doi.org/10.1016/j.cie.2023.109023

[9] Chen, K., Guo, Z., Huang, X., and Jin, Y. (2023, April). LSTM for Return Prediction and Portfolio Optimization in America Stock Market. In FFIT 2022: Proceedings of the International Conference on Financial Innovation, FinTech and Information Technology, FFIT 2022, October 28-30, 2022, Shenzhen, China (p. 305). European Alliance for Innovation.

[10] Thavaneswaran, A., Liang, Y., Paseka, A., Hoque, M. E., & Thulasiram, R. K. (2021, July). A Novel Data Driven Machine Learning Algorithm For Fuzzy Estimates of Optimal Portfolio Weights and Risk Tolerance Coefficient. In 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-6). IEEE.

[11] Choueifaty, Y., and Coignard, Y. “Toward maximum diversification.” The Journal of Portfolio Management, 35(1), 40-51, 2008

[12] Ibanez, F. A. (2023). Diversified Spectral Portfolios: An Unsupervised Learning Approach to Diversification. The Journal of Financial Data Science, Advance online publication. https://doi.org/10.3905/jfds.2023.1.118.

[13] Li, X. (2023). A Regularized K-means Autoregressive Shrinkage Method to Portfolio Optimization. Highlights in Business, Economics and Management, 5, 192–204. https://doi.org/10.54097/hbem.v5i.5076.

[14] Aslam, B., Bhuiyan, R. A., and Zhang, C. (2023). Portfolio Construction with K-Means Clustering Algorithm Based on Three Factors. In MATEC Web Conf. (Vol. 377, Article No. 02006, pp. 1-9). Curtin Global Campus Higher Degree by Research Colloquium (CGCHDRC 2022). DOI: https://doi.org/10.1051/matecconf/202337702006.

[15] Pranata, K. S., Gunawan, A. A. S., and Gaol, F. L. (2023). Development clustering system IDX company with k-means algorithm and DBSCAN based on fundamental indicator and ESG. Procedia Computer Science, 216, 319-327. https://doi.org/10.1016/j.procs.2022.12.142.

[16] Fang, Z., and Chiao, C. “Research on prediction and recommendation of financial stocks based on K-means clustering algorithm optimization.” Journal of Computational Methods in Sciences and Engineering, 21(5), 1081-1089, 2021

[17] Doering, J., Kizys, R., Juan, A. A., Fito, A., and Polat, O. “Metaheuristics for rich portfolio optimization and risk management: Current state and future trends.” Operations Research Perspectives, 6, 100121, 2019.

[18] Thavaneswaran, A., Liang, Y., Yu, N., Paseka, A., and Thulasiram, R. K. “Novel Data-Driven Resilient Portfolio Risk Measures Using Sign and Volatility Correlations.” In 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 1742-1747, IEEE, July 2021.

[19] Bowala, S., and Singh, J. “Optimizing Portfolio Risk of Cryptocurrencies Using Data-Driven Risk Measures.” Journal of Risk and Financial Management, 15(10), 427, 2022.

[20] Chang, C.-C., Lin, Z.-T., Koc, W.-W., Chou, C., & Huang, S.-H. (2016). Affinity Propagation Clustering for Intelligent Portfolio Diversification and Investment Risk Reduction. 2016 7th International Conference on Cloud Computing and Big Data (CCBD), Macau, China, pp. 145-150. doi: 10.1109/CCBD.2016.037.

[21] Kitanovski, D., Mirchev, M., Chorbev, I., & Mishkovski, I. (2022). Cryptocurrency Portfolio Diversification Using Network Community Detection. 2022 30th Telecommunications Forum (TELFOR), Belgrade, Serbia, 1-4. doi: 10.1109/TELFOR56187.2022.9983742.