CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 10 (2) (2025), Pages 1069-1091
p-ISSN: 2086-0382; e-ISSN: 2477-3344

A Systematic Literature Review on Mean-CVaR Based Financial Asset Portfolio Weight Allocation Using K-Means Clustering

Alim Jaizul Wahid[1*], Riaman[2], and Sukono[2]

[1] *Master Program in Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jatinangor, Sumedang 45363, Indonesia*
[2] *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jatinangor, Sumedang 45363, Indonesia*

Abstract

This study examines the integration of the Mean–Conditional Value-at-Risk (Mean-CVaR) model and K-Means Clustering in the allocation financial asset portfolio weight optimization. The growing need for robust and adaptive risk models motivates this review, as existing research often applies CVaR and clustering separately. Using a Systematic Literature Review (SLR) guided by the PRISMA 2020 protocol, data were collected from Scopus, ScienceDirect, and Dimensions databases, yielding 1,598 records. After screening and eligibility verification under academic supervision, six relevant studies were selected for analysis. Bibliometric and qualitative synthesis was performed using RStudio with the Bibliometrix and Biblioshiny packages. The results indicate that CVaR is the dominant risk measure for tail-risk management, while K-Means Clustering is mainly used for asset grouping or scenario generation. Most studies employ metaheuristic solvers such as Genetic Algorithms, Particle Swarm Optimization, or Teaching Learning-Based Optimization, yet direct integration of Mean-CVaR and K-Means within a unified framework remains limited. This review highlights the need for developing hybrid Mean-CVaR and K-Means models applied to multi-asset portfolios, incorporating ESG-adjusted CVaR and machine-learning-based optimization to enhance diversification, sustainability, and resilience across market regimes.

Keywords: Mean-Conditional Value-at-Risk, K-Means Clustering, Portfolio Weight Allocation, Financial Asset Portfolio.

1 Introduction

Capital markets play an important role in the global economy as a means of intermediation between parties with surplus funds and parties who need funding [1].However, market dynamics influenced by price volatility, macroeconomic uncertainty, and interconnections between financial instruments make portfolio management a complex challenge [2], [3]. Since the introduction of Modern Portfolio Theory (MPT) by Markowitz in 1952, the mean–variance approach has become the main framework in portfolio optimization [4], [5]. Although effective in formulating

[*] **Corresponding author**. E-mail: alim24002@mail.unpad.ac.id

Submitted: September 23, 2025 Reviewed: October 07, 2025 Accepted: October 22, 2025
DOI: https://doi.org/10.18860/cauchy.v10i2.36590

the trade-off between return and risk, this model assumes a normal distribution of returns and only uses variance as a measure of risk, making it less responsive to extreme risks (tail risk) [6], [7].

Therefore, to overcome these limitations, Value at Risk (VaR) was introduced as a risk measure that is able to estimate the maximum loss at a certain level of confidence [8].However, VaR does not satisfy the coherence property and does not provide information regarding losses outside the measured quantile [9]. Conditional Value at Risk (CVaR), developed by Rockafellar and Uryasev in 2000, offers a superior solution by focusing on the average loss in the tail of the distribution, as well as fulfilling the coherent nature of risk [9], [10], [11]. Mathematically, the expected return of a portfolio with weights vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and vector expected returns $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ is defined as:

$$E[x] = \boldsymbol{\mu}^T \mathbf{x} \tag{1}$$

Portfolio losses can be represented by the function:

$$f(\mathbf{x}, \mathbf{y}) = -\mathbf{y}^T \mathbf{x} \tag{2}$$

With $\mathbf{y}$ is the realized random return vector of the assets. Let T denote the number of scenarios, z_i the slack variable representing excess loss for scenario i. For a confidence level $\alpha = 0.95$, and γ the Value-at-Risk (VaR) threshold. The feasible set of portfolio weights is denoted by Ω, which typically includes constraints such as $\sum_{j=1}^{n} x_j = 1$ and $x_j \geq 0$ (no short-selling). The Conditional Value at Risk (CVaR) for a fixed portfolio $\mathbf{x}$ is defined as:

$$CVaR_\alpha(\mathbf{x}) = \min_{\gamma, z \geq 0} \left\{ \gamma + \frac{1}{(1-\alpha)T} \sum_{i=1}^{T} z_i \right\} \tag{3}$$

subject to:

$$z_i \geq f(\mathbf{x}, \mathbf{y}_i) - \gamma, \qquad i = 1, 2, \ldots, T \tag{4}$$

Here, the minimization is carried out over γ and the slack variables $\mathbf{z}$, while $\mathbf{x}$ is fixed ensuring that the CVaR definition properly separates the risk measure from the portfolio optimization itself.

Furthermore, the Mean-CVaR model can be formulated with two main approaches, namely:

1. Minimizing CVaR risk for a given minimum return level R:

$$\min_{\substack{\mathbf{x} \in \Omega, \\ \gamma, \mathbf{z} \geq 0}} \left\{ \gamma + \frac{1}{(1-\alpha)T} \sum_{i=1}^{T} z_i \right\} \tag{5}$$

   subject to:

$$\boldsymbol{\mu}^T \mathbf{x} \geq R, \tag{6}$$

$$z_i \geq f(\mathbf{x}, \mathbf{y}_i) - \gamma, \qquad i = 1, 2, \ldots, T \tag{7}$$

2. Solving the trade-off between return and risk with risk aversion parameters λ :

$$\min_{\substack{\mathbf{x} \in \Omega, \\ \gamma, \mathbf{z} \geq 0}} \left\{ -\boldsymbol{\mu}^T \mathbf{x} + \lambda \left(\gamma + \frac{1}{(1-\alpha)T} \sum_{i=1}^{T} z_i \right) \right\} \tag{8}$$

   This formulation clearly distinguishes the CVaR definition (minimized over γ and $\mathbf{z}$ for a fixed portfolio x) from the Mean-CVaR optimization problem (which optimizes over the portfolio weights x). It also confirms that the Mean-CVaR model effectively balances the goal of maximizing expected return and minimizing extreme downside risk, making it more robust and coherent than the traditional Mean-Variance approach [9].

The Mean-CVaR approach combines maximizing expected returns with limiting the risk of extreme losses, and has been widely applied to portfolios of stocks, commodities, and even crypto assets [12], [13], [14]. Previous research has shown that the mean-CVaR approach is more effective in formulating portfolios that adapt to market volatility, such as research conducted by Yu and Liu [15] developed a personalized Mean-CVaR model that adapts portfolio weights to individual risk profiles. Meanwhile, Bedoui et al. [16] integrated Mean-CVaR with the Vine Copula-GARCH-EVT model and genetic algorithm, which demonstrated the effectiveness of CVaR in managing risk on highly volatile assets. Recent research by Jain et al. [17] also emphasized the importance of CVaR through integration with a multi-objective approach based on Teaching Learning-Based Optimization (TLBO) in the context of a sustainable portfolio.

On the other hand, asset grouping using the K-Means Clustering method is widely used in efforts to increase portfolio diversification [18], [19], [20]. As research conducted by Mba and Angaman [21] by applying K-Means in crypto portfolio strategy to differentiate large and small cap assets, while Bulani et al. [22] combining K-Means with Particle Swarm Optimization (PSO) to improve portfolio management of stocks and digital assets. Kaut [23] even using K-Means to generate scenarios within a CVaR-based stochastic programming framework. These studies confirm that K-Means plays a crucial role as a complementary approach in modern portfolio modeling.

Although the topics of mean-CVaR and K-Means clustering have been extensively studied, their integration within a stock portfolio weight allocation framework remains rare. A bibliometric analysis of 1,928 publications from three databases (Scopus, ScienceDirect, Dimensions) for the period 2016–2025 shows that CVaR is a central theme closely related to risk management, optimization, and stochastic programming, while K-Means clustering emerges as a separate cluster focused on asset grouping. No systematic review has been found that maps the trends, methodologies, and research gaps in the integration of these two approaches. Therefore, although Mean-CVaR has proven effective in managing extreme risk and K-Means Clustering plays a crucial role in asset diversification, their integration within a stock portfolio weight allocation framework remains scarce. This suggests a research opportunity to develop hybrid models that are more adaptive and relevant to modern market dynamics.

To address this limitation, the present study aims to provide a more comprehensive understanding of the integration patterns between Mean-CVaR and K-Means in portfolio optimization. By adopting a Systematic Literature Review (SLR) guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework and supported by bibliometric analysis, this review maps the research trends, methodological approaches, and integration directions in this area. It also identifies the diverse applications of K-Means, which commonly serves as a pre-optimization (pre-clustering) stage before the Mean-CVaR optimization process, and examines the range of optimization algorithms used across the reviewed studies.

Accordingly, this study is designed to systematically review and synthesize the existing body of knowledge related to Mean-CVaR-based portfolio optimization and its integration with clustering techniques. Specifically, the objectives of this study are: First, to analyze research trends and bibliometric characteristics of the relevant literature. Second, to classify methodological approaches and optimization models involving Mean-CVaR and K-Means. Third, to develop a conceptual framework that highlights potential pathways for future research in adaptive and diversified portfolio optimization. The overall methodological process of this review follows the PRISMA protocol, as detailed in the next section.

2 Methods

This study employed a Systematic Literature Review (SLR) guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) framework. This approach was used to ensure the literature review process was conducted systematically, transparently, and replicable. The methodological procedure consisted of five main phases: database identification, search and query formulation, screening and eligibility assessment, quality appraisal, and data extraction and synthesis [24], [25]. The literature search was conducted across three major academic databases Scopus, ScienceDirect, and Dimensions, which were selected due to their comprehensive coverage of research in mathematics, computer science, finance, and management. The search was carried out between August 8 and August 15, 2025, ensuring inclusion of the most recent publications. To maintain relevance and quality, the following inclusion parameters were applied:

1) Search with in “Article title, Abstract, Keywords”
2) Articles published in the last 10 years 2016-2025
3) Articles with subject areas "Computer Science; Mathematics; Economics, Econometrics and Finance; Business, Management and Accounting"
4) Document type “article”
5) Articles in English
6) Source type “Journal”
7) Publication stage “Final”
8) articles published open access. The restriction to open-access journals was intentionally applied to ensure transparency and verifiability of the reviewed works. To mitigate potential scope bias, cross-verification was conducted by scanning non-open-access sources, confirming that no unique themes were omitted outside the open-access corpus.

The search strategy uses a combination of keywords that have been classified in Table 1 below.

Table 1: Keyword classification

Type	**Keywords**
A	("Mean-CVaR" OR "Conditional Value-at-Risk" OR "CVaR")
B	("portfolio optimization" OR "portfolio allocation" OR "stock selection")
C	("K-Means clustering" OR "cluster analysis" OR "unsupervised learning")

To operationalize the search, Boolean expressions were formulated by combining the three keyword groups in Table 1. Four Boolean query combinations were obtained, which are summarized in Table 2 :

Table 2: Boolean search query combinations used across three databases

Query Code	Boolean Expression
A	("Mean-CVaR" OR "Conditional Value-at-Risk" OR "CVaR")
A AND B	("Mean-CVaR" OR "Conditional Value-at-Risk" OR "CVaR") AND ("portfolio optimization" OR "portfolio allocation" OR "stock selection")
A AND C	("Mean-CVaR" OR "Conditional Value-at-Risk" OR "CVaR") AND ("K-Means clustering" OR "cluster analysis" OR "unsupervised learning")
A AND B AND C	("Mean-CVaR" OR "Conditional Value-at-Risk" OR "CVaR") AND ("portfolio optimization" OR "portfolio allocation" OR "stock selection") AND ("K-Means clustering" OR "cluster analysis" OR "unsupervised learning")

The number of retrieved records from each query combination is summarized in Table 3.

Table 3: Number of publications from three databases with four types of keywords

Keywords	**Type**	**Scopus***	**Sciencedirect****	**Dimensions*****	**Total**
Keyword 1	A	771	272	318	1,361
Keyword 2	A AND B	89	63	47	199
Keyword 3	A AND C	5	26	3	34
Keyword 4	A AND B AND C	0	4	0	4
Total		865	365	368	1,598

*sourced from https://www.scopus.com/.
**sourced from https://www.sciencedirect.com/.
***sourced from https://www.dimensions.ai/.

The selection process adhered to the PRISMA 2020 protocol, ensuring transparency, reproducibility, and methodological rigor throughout the systematic review. A total of 1,598 records were initially identified across three academic databases, consisting of 865 from Scopus, 365 from ScienceDirect, and 368 from Dimensions. Duplicate entries were automatically detected and removed using RStudio (version 4.3.0), which eliminated 243 duplicates and exclusion of ineligible records by automation, 1,146 unique documents remained for the initial screening phase.

The title and abstract screening was performed using Mendeley Reference Manager, which served as the main reference organization and coding tool. Each article was categorized based on its thematic relevance to the research scope [25]. During this phase, studies that did not address any of the core elements namely Mean-CVaR, portfolio optimization, or clustering methods were excluded. This step removed 782 papers, leaving 531 articles were retained for full-text assessment.

In the full-text eligibility screening stage, each study was examined in detail to assess (1) the methodological transparency of its portfolio optimization framework, (2) the explicit application of Mean-CVaR as a risk model, and (3) the incorporation of K-Means clustering or similar unsupervised learning techniques for asset grouping or diversification. Articles employing alternative risk measures (e.g., Value-at-Risk or Mean-Variance), lacking any mathematical optimization component, or focusing on non-financial portfolio contexts were excluded from further analysis.

To maintain objectivity and ensure the reliability of the inclusion process, the screening and eligibility assessments were conducted by the first author under the academic supervision of the second and third authors. Every inclusion or exclusion decision was reviewed through iterative academic consultations until a complete consensus was achieved. Although the review process was primarily led by a single researcher, this supervisory verification mechanism provided sufficient methodological reliability and minimized the potential for subjective selection bias.

The inclusion criteria were carefully defined to capture two complementary research domains that have strong integration potential: (1) Mean-CVaR based risk optimization and (2) K-Means-based asset grouping. Articles were included if they satisfied one or more of the following conditions The inclusion criteria were carefully defined to capture two complementary research domains with high integration potential Mean-CVaR based risk optimization and K-Means-based asset grouping. Studies were considered eligible if they were published between 2016 and 2025 in peer-reviewed English-language journals, were available in full-text open-access format, and explicitly addressed Mean-CVaR or Conditional Value-at-Risk (CVaR) within the context of portfolio optimization. In addition, papers that applied K-Means clustering or other unsupervised learning techniques as part of asset grouping, portfolio construction, or diversification strategies were also included. Conversely, studies were excluded if they did not discuss or contribute to portfolio optimization, failed to employ either Mean-CVaR or K-Means clustering, or focused solely on alternative risk models such as Value-at-Risk or Mean-Variance without conceptual or methodological relevance to Mean-CVaR. Articles that were not accessible in full-text form or were restricted behind paywalls were likewise excluded from the review.

Applying these systematic inclusion and exclusion criteria resulted in a final dataset of 6 eligible studies that fully met the established methodological standards. These selected papers collectively represent the current state of research at the intersection of Mean-CVaR based optimization and K-Means-based asset grouping. Together, they form the conceptual and analytical foundation synthesized and discussed in the subsequent section of this study. The overall selection flow, including the number of articles identified, screened, excluded, and ultimately included, is illustrated in Fig. 1 (PRISMA flow diagram), which summarizes the screening and eligibility outcomes at each stage of the review process.

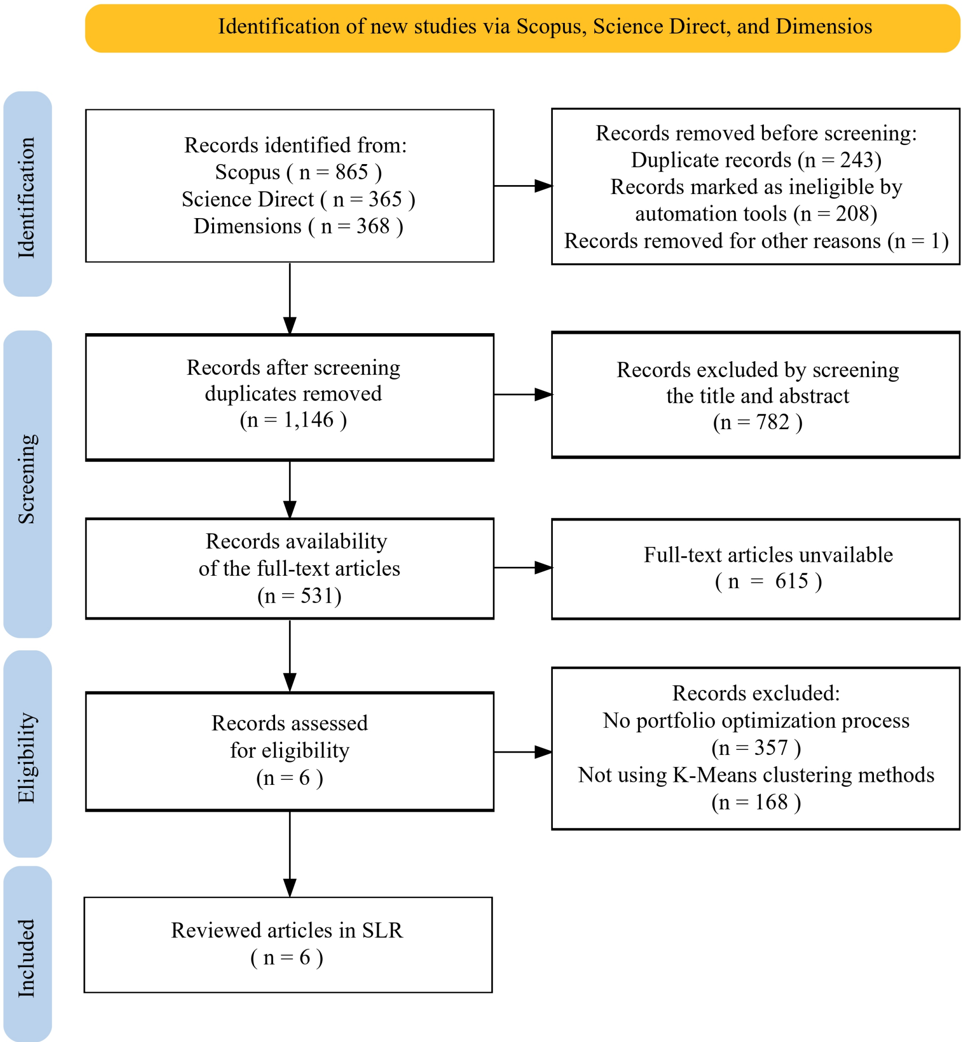


Figure 1: PRISMA Diagram

Fig. 1 presents the PRISMA 2020 flow diagram summarizing the identification, screening, eligibility, and inclusion stages. The diagram also specifies records removed before screening, duplicates, unavailable full-texts, and final inclusion count, the selection results based on the relevance of the title and abstract of 1,146 articles are obtained, referred to as Dataset 1, which will be used for bibliometric analysis. After that, a full text screening process was carried out with the acquisition of 531 articles. Based on the results of the full text selection, eligible articles were selected discussing the portfolio optimization process and the K-Means clustering method, resulting in 6 articles also referred to as Dataset 2.

Although the final dataset consisted of only six eligible studies, this number is sufficient and methodologically justified within the context of this SLR. The relatively small corpus resulted from the combination of two highly specific criteria: (1) the explicit integration of Mean-CVaR based portfolio K-Means-based asset clustering, and (2) the inclusion of only peer-reviewed and open-access journal articles to ensure transparency and reproducibility. These six studies were not chosen merely based on availability, but because they collectively represent the entire methodological spectrum identified in the literstochastic programming, mixed-integer optimization, hybrid metaheuristic, and clustering-based frameworks. The thematic and mathematical diversity among these papers ensures that the synthesis remains representative of the current state of research despite the limited number of sources. Hence, maintaining a focused corpus of six high-quality, conceptually integrated studies.

This Dataset 2 will be used as the initial state of the art which will be analyzed further in the results and discussion sections. After final article selection, bibliometric and qualitative analyses were conducted using RStudio with the Biblioshiny package. The bibliometric analysis visualized co-occurrence networks, thematic maps, and productivity trends, while qualitative synthesis examined the methodological convergence of the six studies. Together, these analyses formed the foundation for the conceptual framework presented in the Results and Discussion Section 3.

3 Results and Discussion

This section integrates the findings from the bibliometric analysis and the systematic literature review (SLR). The bibliometric results reveal how the research domain surrounding Mean-CVaR and K-Means clustering has evolved, while the SLR component explores six representative studies that illustrate the conceptual and methodological patterns of integration between these two approaches. Together, these analyses provide not only a descriptive mapping of the field but also an interpretive understanding of the intellectual structure and emerging trends.

3.1 Bibliometric Maps Results Using RStudio Software Procedure

The bibliometric research procedure in this study is systematically illustrated through the workflow shown in Fig. 2. This workflow provides a comprehensive overview of the sequential stages conducted in the bibliometric analysis. The process begins with defining the research domain, which in this case focuses on the integration of the Mean–CVaR model and K-Means clustering in financial portfolio optimization. After establishing the scope, relevant data were collected from three major academic databases Scopus, ScienceDirect, and Dimensions to ensure comprehensive coverage of high-quality and peer-reviewed studies.

Subsequently, bibliometric data mining was carried out, which included data cleaning, normalization of author and keyword metadata, and removal of duplicate records. The cleaned dataset was then analyzed using RStudio with the Bibliometrix and Biblioshiny packages to extract quantitative indicators such as publication trends, citation patterns, co-authorship networks, keyword co-occurrence, and thematic evolution.

Finally, the resulting outputs were visualized and interpreted to map the current state-of-the-art in this research area. These mappings enabled the identification of major research clusters, methodological developments, and thematic gaps, thereby providing a solid foundation for recognizing emerging trends and future research opportunities in hybrid portfolio optimization frameworks.

As illustrated in Fig. 2, the bibliometric procedure began by defining the research scope Mean-CVaR Based Financial Asset Portfolio Weight Allocation Using K-Means Clustering. A systematic search was then performed across three major databases (Scopus, ScienceDirect, and Dimensions) to ensure comprehensive coverage of relevant studies. The bibliometric data mining phase included executing predefined search criteria, reviewing the retrieved records for relevance, and exporting the selected datasets. These records were subsequently imported into RStudio and processed using the Bibliometrix package to extract key analytical parameters such as author productivity, publication sources, article citations, keyword occurrences, and country-level contributions.

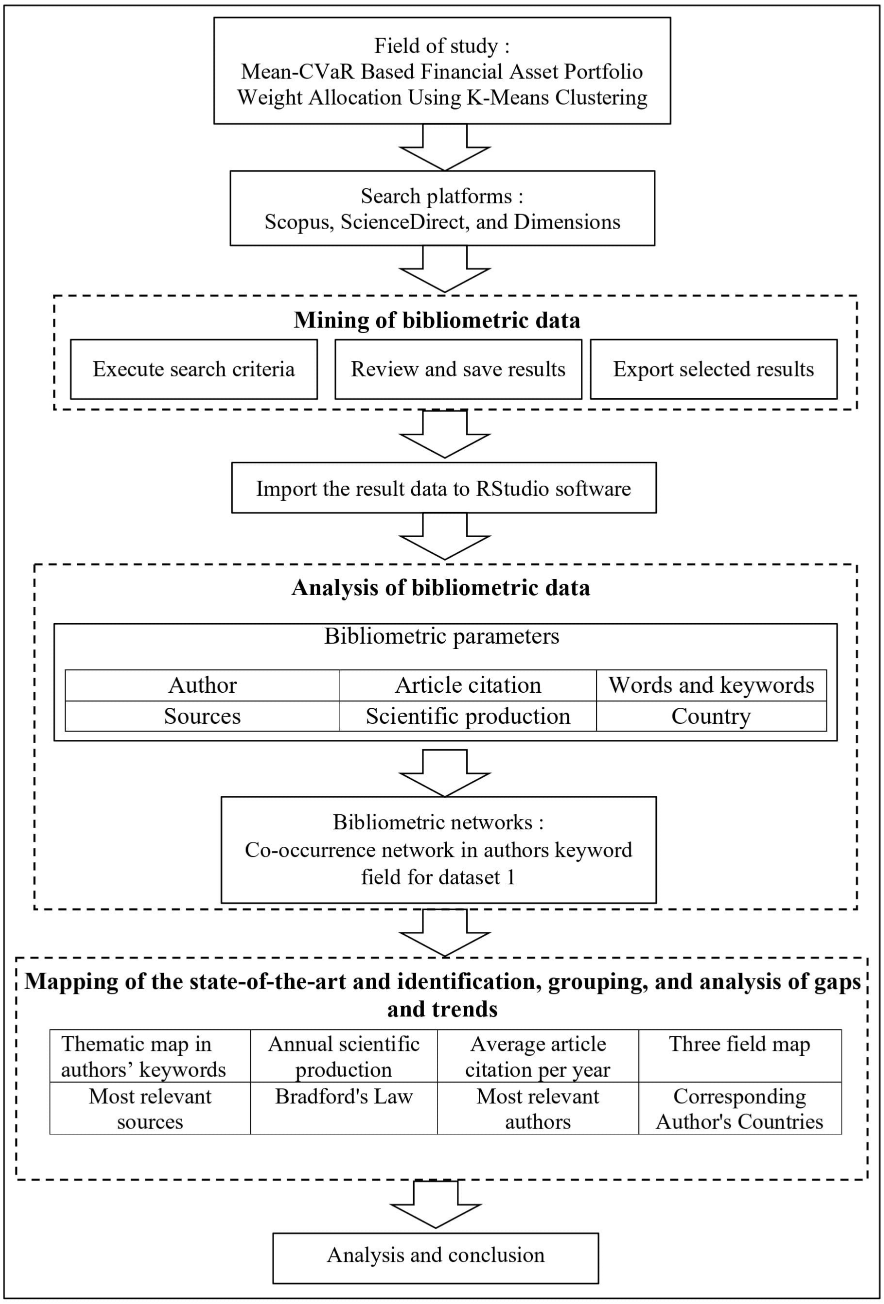


Figure 2: Bibliometric Mapping Output Diagram

The following phase involved network-based analyses, particularly the co-occurrence mapping of author keywords to uncover thematic relationships and intellectual structures within the field. These network outputs were complemented by higher-order analyses including thematic evolution, annual scientific production, citation dynamics, Bradford's Law of source dispersion, and author-country collaboration networks. This comprehensive analytical workflow not only visualizes the bibliometric landscape but also enables the interpretation of underlying research patterns, the identification of emerging themes, and the detection of gaps that guide future investigations in hybrid portfolio optimization research.

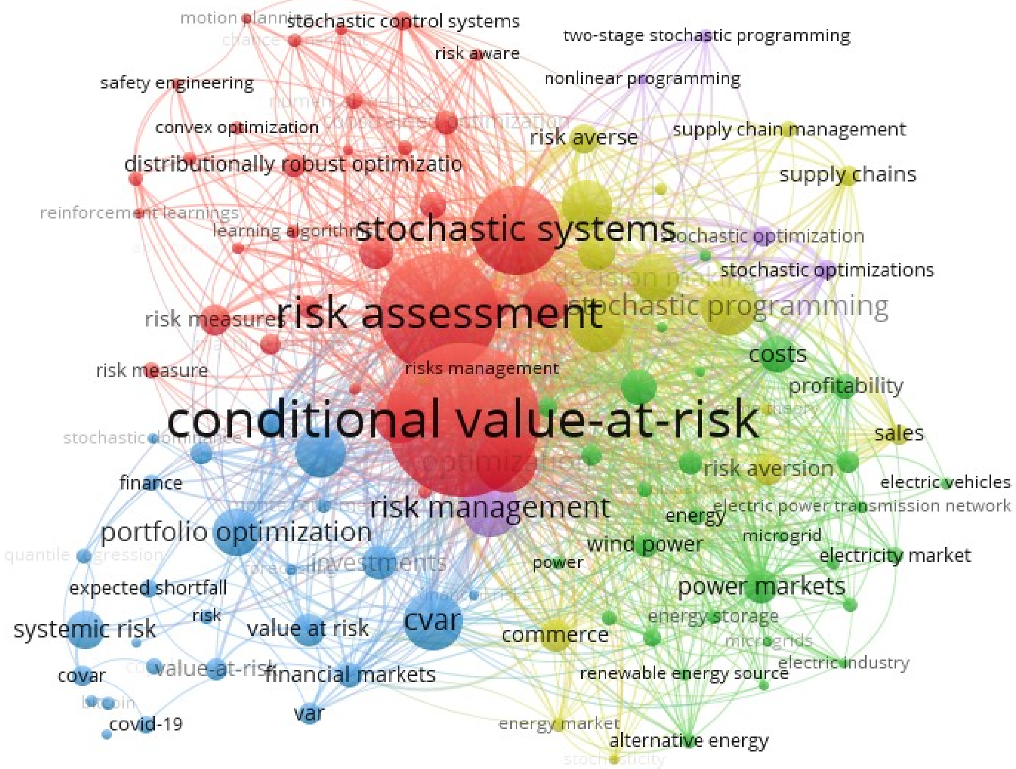


Figure 3: Co-occurrence Network

The co-occurrence network in Fig. 3 illustrates the conceptual landscape of research integrating ris, optimization, and clustering techniques. The visualization highlights "Conditional Value-at-Risk (CVaR)" as the"CVaR" functions as the nexus connecting multiple research streams—from quantitative optimization to resilience and sustainability indicating that the measure of tail-risk has evolved into a universal anchor for financial and operational decision-making across diverse contexts.

The blue cl, encompassing keywords such as "risk management", "portfolio optimization", and "systemic risk", represents the traditional foundation of financial risk modeling. Its prominence suggests that the CVaR framework remains the preferred model for handling extreme losses, particularly in stock and asset allocation problems. The high density of interconnections within this cluster indicates that current research continues to refine the mathematical rigor and applicability of CVaR-based optimization, rather than diverging from it.

The red cluster, containing "stochastic programming", "decision making", and "optimization", reflects the methodological evolution ofCVaR core underscores that optimization under uncertainty has become a key direction in advancing portfolio theory. Future studies may build upon this link by developing hybrid techniques that combine CVaR with heuristic and metaheuristic algorithms.

Meanwhile, the green cluster connects "sustainability", "energy markets", and "supply chain management", illustrating a cross-disciplinary expansion of CVaR applications beyond the financial domain. The integration of CVaR into sustainability and energy-related contexts reflects a shift from purely financial risk toward systemic and environmental risk assessment particularly in renewable eCVaR's adaptability in evaluating ESG (Environmental, Social, and Governance) risks.

Finally, the purple cluster, consisting of "COVID-19", "machine learning", and "energy", captures emerging analytical frontiers. These topics reveal how global disruptions and technological innovations are driving CVaR modeling toward dynamic, AI-enhanced frameworks. The frequent co-occurrence of "machine learning" with risk-related terms suggests a transition toward data-driven portfolio clustering and adaptive risk estimation potentially integrating "K-Means" and deep-learning architectures.

Overall, the co-occurrence network demonstrates a paradigm shift from static, finance-centered risk optimization toward dynamic, multi-domain risk modeling enhanced by artificial intelligence and sustainability principles. This evolution not only broadens the application of CVaR but also underscores the need for future hybrid models that combine Mean–CVaR optimiza, K-Means clustering, and machine-learning-based scenario generation to address increasingly complex and volatile market environments.

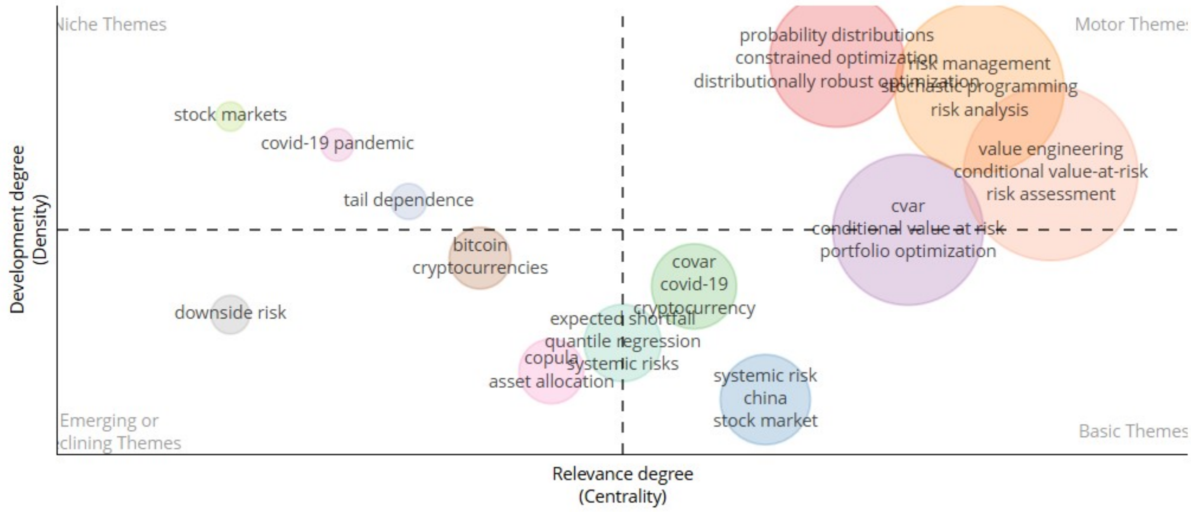


Figure 4: Thematic Map

Fig. 4, presents the thematic map generated from Dataset 2, illustrating the structural evolution of research topics according to their development degree (density) and relevance degree (centrality). The four quadrants represent distinct thematic statuses motor themes, niche themes, basic themes, and emerging or declining themes which collectively reveal how the field of CVaR-based portfolio optimization has matured and diversified.

1) Upper Right Quadrant (Motor Themes)
   comprises topics such as "conditional value-at-risk", "risk assessment", "risk management", and "stochastic programming". These are well-developed and central themes, forming the methodological backbone of this research domain. Their strong centrality indicates that CVaR remains the core paradigm for quantitative risk modeling, while their high density reflects continuous methodological refinement through integration with advanced optimization frameworks such as distributionally robust and stochastic programming. This suggests that CVaR-based modeling is not static but continuously evolving to address uncertainty, particularly in dynamic financial and energy markets.
2) Upper Left Quadrant (Niche Themes) (Niche Themes)
   contains specialized areas including "probability distributions", "robust optimization", and "value engineering", which exhibit strong internal coherence but limited connectivity with broader research streams. These topics represent methodological extensions focused on model generalization and reliability analysis indicating a movement toward formalizing risk estimation and control under extreme or rare events. Their niche position implies potential growth into core themes if future research succeeds in operationalizing these methods for practical portfolio optimization.
3) Lower Right Quadrant (Basic Themes)
   terms such as "CVaR", "portfolio optimization", and "systemic risk" occupy a foundational position. Despite moderate density, their high centrality underscores their fundamental

role as building blocks linking theoretical development with applied modeling. These topics serve as conceptual anchors for new hybrid approaches that merge Mean-CVaR optimization with machine learning or clustering algorithms like K-Means, pointing toward an integrative trend in future research.

4) Lower Left Quadrant (Emerging or Declining Themes)
includes "cryptocurrencies", "bitcoin", and "downside risk", representing nascent or context-specific research areas. Their low centrality and density indicate that these topics are still exploratory but potentially promising, especially given the increasing attention to digital asset volatility and tail-risk modeling. Future research may expand these themes by embedding CVaR based optimization within cryptocurrency portfolio frameworks or by exploring clustering-based risk segmentation in high-frequency data environments.

Overall, the thematic map suggests a maturing but evolving research field. Core risk modeling frameworks such as CVaR and stochastic programming continue to dominate, while peripheral topics like robust optimization and cryptocurrency risk represent the next frontier for theoretical and computational innovation. The observed structure implies that future studies should move beyond isolated applications and develop integrated hybrid frameworks combining Mean-CVaR, K-Means, and data-driven optimization to enhance both interpretability and adaptivity in modern portfolio decision-making.

3.2 Evolution Analyzes with tools package Biblioshiny

When uploading the combined file of the three databases to biblioshiny, the following analysis results were obtained.

Figure 5: Main Information

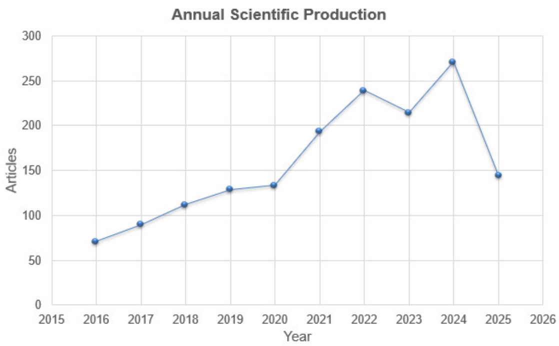
Fig. 5 summarizes the bibliometric characteristics of the merged dataset compiled from Scopus, ScienceDirect, and Dimensions databases for the period 2016–2025. A total of 1,598 documents from 605 journal sources were analyzed, reflecting a substantial and diversified body of research at the intersection of Conditional Value-at-Risk (CVaR), portfolio optimization, and clustering-based analytical modeling. The annual publication growth of 8.17% indicates a consistent and sustained development of this research domain, suggesting that CVaR remains an essential risk metric that continues to attract attention due to its applicability in both financial and operational optimization contexts.

The bibliometric indicators show a strong collaborative culture within this research community. Of the 4,197 contributing authors, only 141 are single-author papers, while the average number of 3.22 co-authors per document signifies that most studies rely on interdisciplinary cooperation, particularly between mathematicians, financial analysts, and computer scientists. Notably, 11.64% of the works involve international co-authorship, suggesting that research on CVaR-based modeling has started to transcend regional boundaries and evolve into a globally connected scientific network. This growth in collaboration likely stems from the computational nature of the field, which often requires access to large datasets and specialized optimization expertise distributed across research centers.

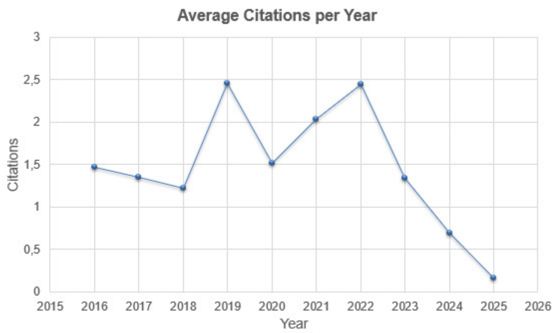
The average publication age of 3.62 years demonstrates that this is a young and actively developing research area, where new frameworks and algorithms are frequently proposed. The mean citation count of 7.645 citations per document suggests moderate but consistent academic impact, reflecting the methodological rather than empirical focus of many studies in this area—where new mathematical formulations, optimization procedures, or hybrid models are proposed rather than large-scale empirical tests.

The dataset contains 36,297 references and 2,602 distinct author keywords, indicating both conceptual richness and methodological diversity. The extensive reference network shows that this domain builds upon a wide range of disciplines, from operations research and stochastic optimization to machine learning and financial econometrics. This interconnectedness highlights how the field of Mean–CVaR based portfolio optimization is evolving toward more integrated, data-intensive paradigms that combine theoretical rigor with computational adaptability.

Taken together, these bibliometric patterns reveal that the study of CVaR and portfolio optimization has entered a phase of methodological convergence where risk modeling, clustering, and optimization are increasingly viewed as complementary rather than isolated techniques. Future research is expected to expand these collaborations further, emphasizing hybrid optimization frameworks that leverage K-Means clustering and other unsupervised learning methods to improve the stability, interpretability, and scalability of risk-based portfolio allocation models.



(a) Annual Scientific Production



(b) Average Citations per Year

Figure 6: Annual Trends in Publications and Citations

Based on Fig. 6a, shows the trend of Annual Scientific Production from 2016 to 2025, which reflects the developmental trajectory of research integrating Conditional Value-at-Risk (CVaR), optimization, and clustering methods. The number of publications increased steadily during the early stage (2016–2020), rising from 71 to 134 articles. This phase represents the foundational period when CVaR-based portfolio optimization began to gain broader acceptance as a standard model for risk-sensitive decision making. A sharp increase occurred after 2020, peaking at 239 publications in 2022 and 271 in 2024, signaling a strong research expansion. This acceleration coincides with the global surge in data-driven financial analytics and the growing integration of machine learning techniques such as K-Means clustering and metaheuristic optimization into portfolio modeling frameworks. The temporary decline observed in 2025 is likely due to incomplete indexing of recent publications rather than an actual drop in productivity a common pattern in bibliometric analyses when the current year's data are still accumulating. Overall, the consistent growth and high publication volume over the last five years indicate that this field has reached a mature but still expanding phase, driven by interdisciplinary collaboration between mathematics, computer science, and finance.

On the other hand, Fig. 6b presents the trend of Average Citations per Year, reflecting the academic influence and visibility of published works. The citation pattern does not perfectly align with publication volume: the highest average citations occurred in 2019 (2.45) and 2022 (2.44). These peaks suggest that papers published during these years were particularly influential likely those proposing new formulations of Mean–CVaR optimization, or applying CVaR to emerging contexts such as energy systems, cryptocurrency portfolios, and machine-learning-based risk estimation. After 2022, the average citation rate shows a gradual decline, which is typical in time-dependent citation patterns since more recent works (2023–2025) have had limited exposure time to accumulate citations. This time-lag effect emphasizes that rapid growth in publication volume does not immediately translate into citation impact. Instead, influence builds over time as the community validates, adopts, and extends novel methods.

Taken together, Fig. 6 reveal a quantitative expansion followed by qualitative consolidation in the field. While publication output continues to rise, citation peaks correspond to methodological breakthroughs rather than sheer volume. This dynamic suggests that future research impact will depend less on the number of studies and more on their conceptual novelty for example, developing hybrid Mean-CVaR and K-Means models, or combining risk optimization with AI-based clustering and scenario generation to address increasingly complex financial and operational uncertainties.

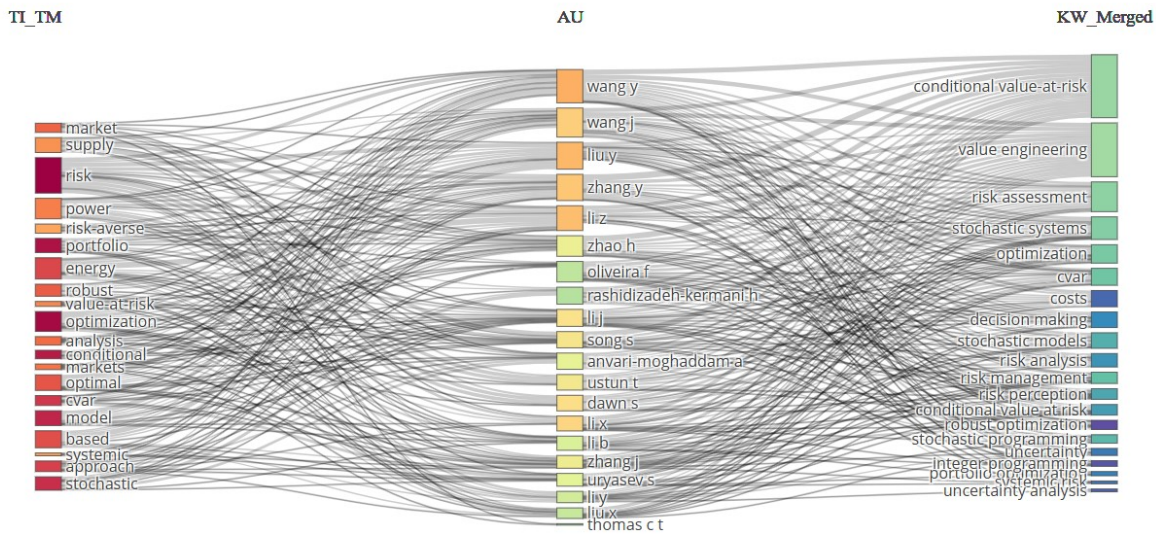


Figure 7: Three plot fields Title, Author, and Keyword

Fig. 7 illustrates the three-field plot linking the most frequent research topics (TI_{TM}), authors (AU), and merged keywords (KW_{Merged}). This visualization reveals the intellectual structure and thematic coherence within the research domain of Conditional Value-at-Risk (CVaR), optimization, and risk-based portfolio modeling. The connections across the three dimensions demonstrate how specific authors contribute to shaping methodological and conceptual directions in this field. On the left, frequently occurring title terms such as “risk,” “portfolio,” “optimization,” “stochastic,” “power,” and “energy” indicate the dominant research contexts and methodological focus. These terms reflect a strong orientation toward quantitative modeling and uncertainty management consistent with the mathematical and operational foundations of Mean–CVaR based portfolio optimization. At the center, several prolific authors notably Wang Y, Zhao H, Liu Y, and Anvari-Moghaddam A serve as major connectors within the research network. Their works frequently appear at the intersection of risk modeling, stochastic optimization, and energy applications, suggesting that they are central figures driving interdisciplinary research that extends CVaR beyond finance into domains such as power systems and renewable energy management. This pattern also highlights that the evolution of CVaR research has been shaped by authors who bridge mathematical optimization with real-world applications in energy and sustainability. On the right, recurrent keywords such as "conditional value-at-risk," "risk assessment," "stochastic systems," "optimization," and "robust optimization" demonstrate strong methodological coherence across studies. The repeated pairing of these terms with "decision-making" and "risk management" suggests that the field has matured into a structured discipline that emphasizes quantitative decision support under uncertainty. The appearance of "value engineering" and "uncertainty analysis" further signals a trend toward integrating engineering and computational perspectives into financial and operational risk frameworks.

Collectively, the three-field plot indicates a tightly interconnected research ecosystem in which methodological innovation (keywords) aligns closely with the thematic direction (titles) and the academic contributors (authors). This structure demonstrates both intellectual consolidation and specialization a sign of a mature but evolving field. The clustering of high-frequency terms around CVaR and optimization also suggests that future developments are likely to focus on hybrid integration, combining Mean-CVaR optimization with machine learning and clustering techniques such as K-Means, to achieve more adaptive and data-driven portfolio allocation frameworks.

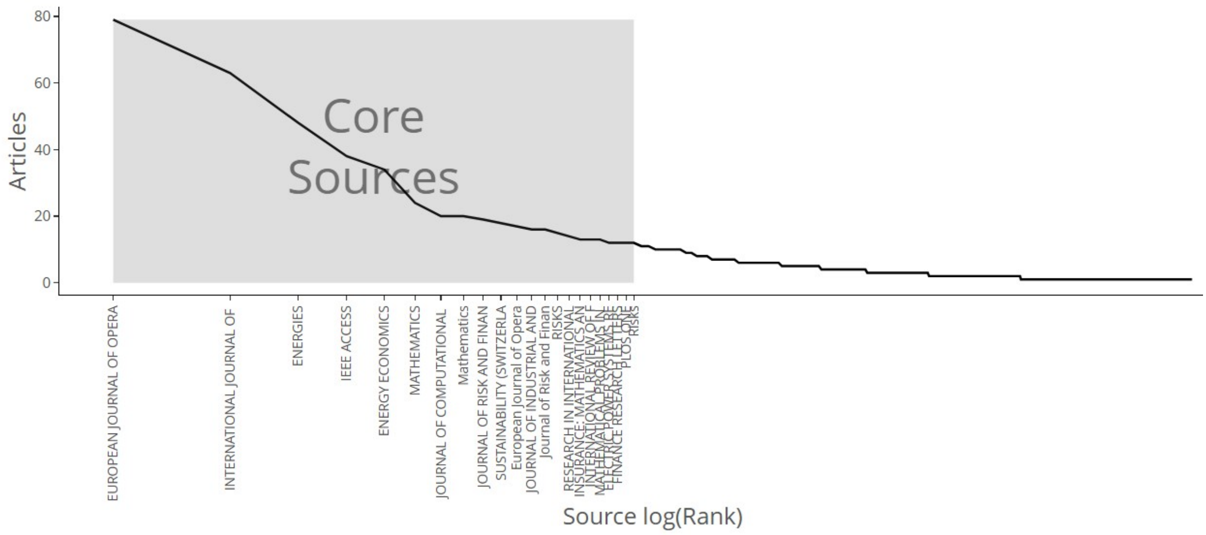
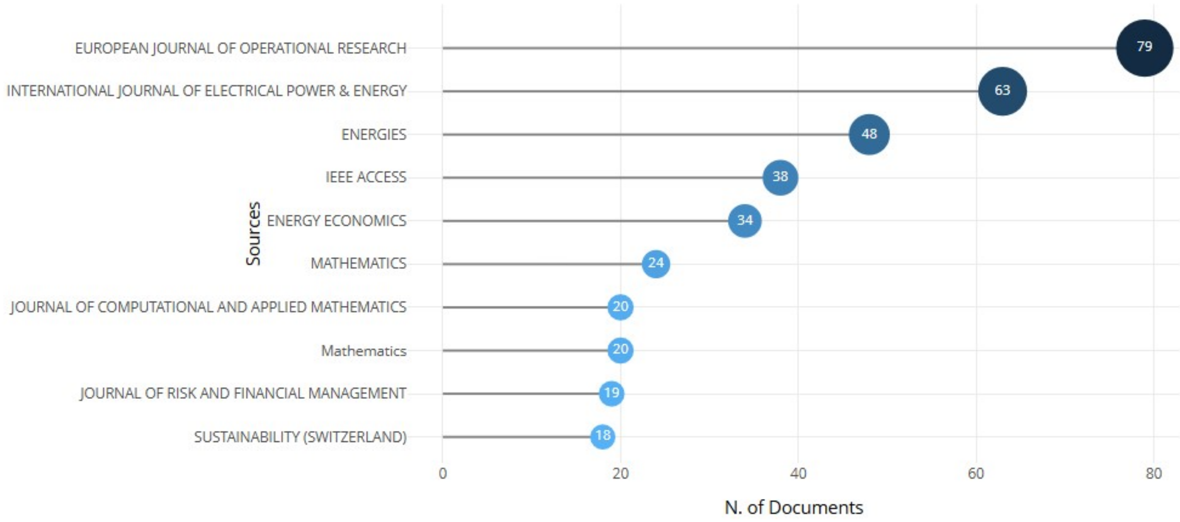


Figure 8: Core Sources Based on Bradford's Law

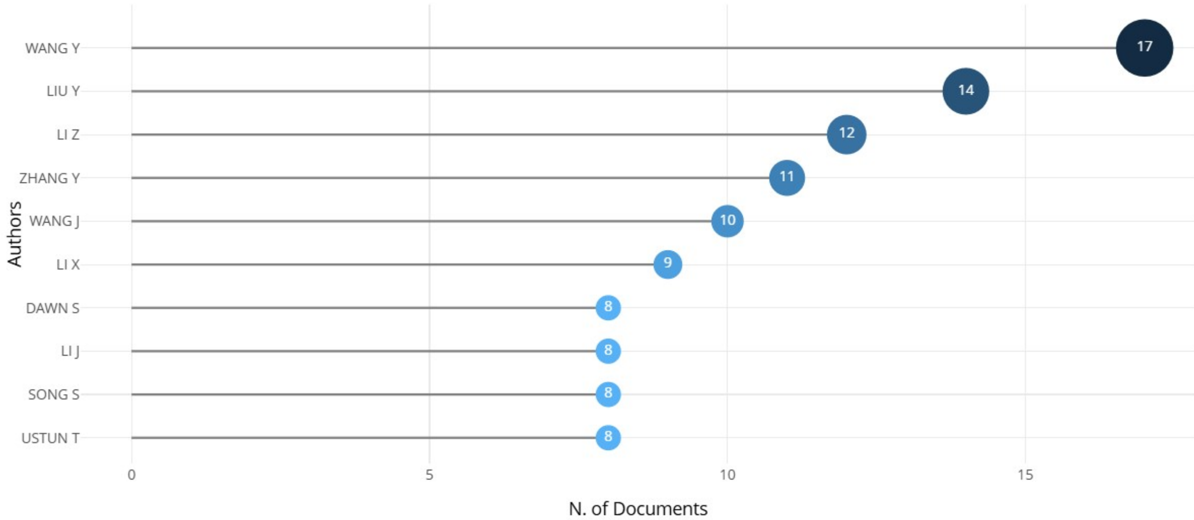
Fig. 8 illustrates the distribution of journals according to Bradford's Law, which divides publication sources into three zones based on their productivity and influence. Zone 1 (core zone) comprises 27 journals that published 533 documents, including Sustainability (Switzerland), IEEE Access, Energy Reports, Energies, and Journal of Cleaner Production. These journals represent the central platform for research integrating Conditional Value-at-Risk (CVaR), portfolio optimization, and clustering-based modeling, reflecting the field's strong methodological and interdisciplinary orientation toward finance, energy, and sustainability.

Zone 2 (100 sources; 526 documents) and Zone 3 (478 sources; 539 documents) show a wider dispersion of publications across numerous journals, indicating knowledge diffusion beyond the core outlets. This pattern suggests that while a few journals dominate in impact, new and peripheral journals are actively contributing to the diversification of approaches, particularly those linking CVaR optimization with machine learning and unsupervised clustering methods. Overall, the distribution confirms a dual trend in the literature: consolidation around high-impact journals that define the theoretical foundation, and simultaneous expansion through diverse outlets exploring hybrid and application-oriented research. This dynamic reflects a mature but continuously evolving research ecosystem.

Fig. 9a shows that Sustainability (Switzerland), IEEE Access, and Energy Reports are the dominant publication sources, reflecting a strong intersection between risk modeling, optimization, and sustainable finance. Although journals like Applied Energy and Journal of Cleaner Production



(a) Most Relevant Source



(b) Most Relevant Author

Figure 9: Core Publication Sources and Influential Authors

publish fewer articles, their high citation rates highlight the emphasis on methodological depth and real-world applicability over quantity. Fig. 9b reveals that Wang Y, Zhang Y, and Zhao H are the most influential authors, consistently advancing stochastic and optimization-based CVaR research. Their recurring contributions across top journals indicate a central research network driving theoretical refinement and cross-disciplinary applications. Overall, both figures illustrate that the field is shaped by a concentrated core of journals and scholars who sustain its methodological and thematic development.

Fig. 10 shows that research output is dominated by China (661 documents), followed by Malaysia (141) and India (111). This concentration reflects Asia's growing leadership in quantitative finance and optimization research, driven by strong academic investment and technological innovation. The notable contributions from Saudi Arabia, Pakistan, and Iran further indicate the expansion of this field across the Middle East and South Asia, highlighting a shared focus on risk modeling and energy-oriented optimization. Overall, the pattern demonstrates that the development of Mean–CVaR and clustering-based portfolio research is centered in Asia, where collaborative and application-driven studies continue to strengthen the global relevance of this domain.

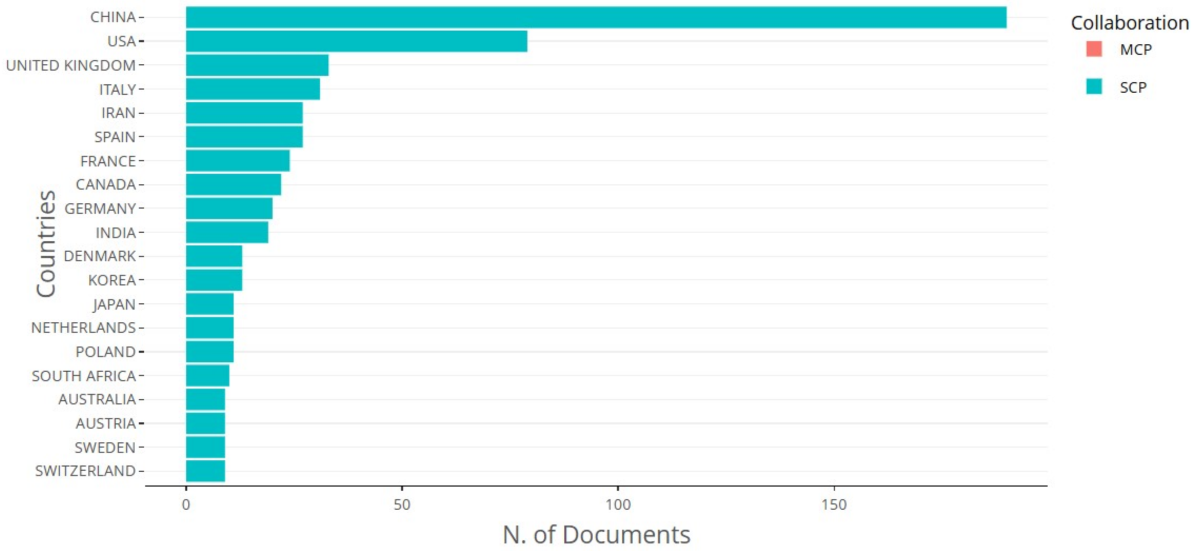


Figure 10: Corresponding Author's Countries

3.3 The Results of Systematic Reviews

This section presents the results of a literature review of Dataset 2, a database containing six articles selected in the final stage. The articles in Dataset 2 were published between 2021 and 2025. The aim of this review is to identify how Mean–CVaR has been applied to portfolio weight allocation and to what extent K-Means clustering has been integrated as a supporting or complementary methodology in financial asset optimization.

1) Systematic Literature Review (SLR) Results
   An initial analysis was conducted to examine the use of CVaR, portfolio optimization methods, and the application of K-Means Clustering in grouping assets within the portfolio weight allocation framework. The mapping results can be seen in Table 4.

Table 4: Relevant Articles SLR Results

Author	CVaR	Portfolio Weight Allocation	Portfolio Optimization	K-Means Clustering
Bedoui et al. (2023)[16]	✓	✓	✓	-
Mba and Angaman (2023)[21]	-	✓	✓	✓
Yu and Liu (2021)[15]	✓	✓	✓	-
Bulani et al. (2025)[22]	✓	-	✓	✓
Kaut (2021)[23]	✓	-	✓	✓
Jain et al. (2025)[17]	✓	-	✓	-
Our research	✓	✓	✓	✓

   Based on Table 4, it can be seen that most studies use CVaR as the main risk measure in portfolio optimization, for example the studies of Bedoui et al. [16] and Jain et al. [17] combine CVaR with advanced optimization techniques such as Genetic Algorithm, Particle Swarm Optimization (PSO), or Teaching Learning-Based Optimization (TLBO). However, direct integration with K-Means Clustering is still limited, namely only Mba and Angaman [21], Bulani et al. [22] and Kaut [23] which utilizes it in the pre-clustering or scenario generation stage. Other research was conducted by Yu and Liu [15] more emphasis is placed on metaheuristic-based optimization or personalized risk weighting. Thus, research trends indicate the dominance of CVaR as a risk measure, but the integration of CVaR and K-Means remains a relatively open area for exploration.

2) Research Topic Methods and Applications

This section aims to answer the first research question, namely: What methods have been used by previous researchers to solve portfolio weight allocation problems using Mean-CVaR and K-Means Clustering?. The summary of methods and their applications is presented in Table 5.

Table 5: Methods and Applications in Related Studies

Author	**Research purposes**	**Method**	**Application/Case**
Bedoui et al. (2023)[16]	Multi-asset portfolio optimization with Vine Copula-GARCH-EVT based CVaR model	NSGA-II Genetic Algorithm, Monte Carlo Simulation	Bitcoin, Gold, Oil, Stock Index
Mba and Angaman (2023)[21]	Examining the benefits of large & small crypto diversification using K-Means & Entropy Pooling	K-Means Classification, Entropy Pooling, normal & skew-t distribution	14 cryptocurrencies during the 2022 crash
Yu and Liu (2021)[15]	Personalized portfolio recommendations based on risk tolerance	Fuzzy Comprehensive Evaluation, Copula-GARCH, Monte Carlo, PSO	Stocks, bonds, mutual funds
Bulani et al. (2025)[22]	Improving portfolio management with clustering & PSO	K-Means, PSO, rebalancing	Stocks & cryptocurrencies 2015–2023
Kaut (2021)[23]	Scenario generation from historical data for portfolio optimization	K-Means, MIP moment matching, Wasserstein distance	Market data for the CVaR stochastic model
Jain et al. (2025)[17]	Multi-objective portfolio optimization based on ESG & downside risk	REM Clustering + TODIM, BN-XGBoost, TLBO	Nifty 100 stocks, ESG constraints

Based on Table 5, it can be seen that each study has a different orientation, for example Bedoui et al. [16] emphasizes multi-asset tail risk management, Mba and Angaman [21] highlighting the benefits of crypto diversification, Yu and Liu [15] introducing personalized portfolios, Bulani et al. [22] focus on the combination of clustering and metaheuristic optimization, Kaut [23] emphasize scenario generation, whereas Jain et al. [17] integrating ESG factors into the optimization framework. This pattern demonstrates a wide variety of methods, but the use of K-Means is still primarily focused on asset preprocessing rather than as a core component of portfolio weight allocation.

3) Variable Specifications and Risk Measurement Approaches
The primary variables and risk measurement approaches used in the selected studies were asset returns (stocks, crypto, bonds, indices, or multi-asset combinations). CVaR is the most commonly used risk measure, with some studies combining it with other methods such as Extreme Value Theory, Monte Carlo Simulation, or alternative risk measures like the Sharpe Ratio and Semi-Variance. This demonstrates a strong focus on tail risk management in portfolio optimization.

From Table 6 it can be seen that although CVaR is the dominant risk measure, several studies enrich the model with additional measures such as the Sharpe Ratio used in Bulani et al. [22]'s study, then VaR was used in Mba and Angaman [21] 's research, or higher moments such as skewness and kurtosis are used by Jain et al. [17] .This signals a shift in trend from a focus solely on tail risk to a more comprehensive risk measurement model, including the sustainability (ESG) dimension.

4) Comparative Summary of Reviewed Studies

Table 6: Main Variables and Risk Measurement Approaches

Author	Main Variables	Risk Measurement Approach	Special Notes
Bedoui et al. (2023)[16]	Asset returns (BTC, Gold, Oil, stock indices)	EVT + Vine Copula-GARCH based CVaR	Tail risk focus, multi-objective GA optimization
Mba and Angaman (2023)[21]	Return crypto large & small cap	VaR, Entropy Pooling	K-Means inter-cluster diversification
Yu and Liu (2021)[15]	Financial asset returns & investor risk profiles	CVaR from Copula-GARCH + Monte Carlo	Personalize portfolio with FCE
Bulani et al. (2025)[22]	Stock & crypto returns	Sharpe, Adjusted Sharpe, Sortino	Clustering for dynamic diversification
Kaut (2021)[23]	Historical return data	CVaR in stochastic models	Scenario selection & data consistency
Jain et al. (2025)[17]	Nifty 100 stock returns + ESG score	CVaR, Semi-Variance, Skewness, Kurtosis	ESG integration and diversification (Gini-Simpson)

To strengthen methodological synthesis and evidence reliability, a comparative framework was constructed, as shown in Table 6, summarizing datasets, solvers, validation metrics, and study limitations.

Table 7: Comparative Summary of Reviewed Studies

Author	Asset Class	Period	Risk Measure (α)	Constraints (Ω)	Solver / Algorithm	Validation Metrics
Bedoui et al. (2023)[16]	Bitcoin, Gold, Oil, Indices	2017–2022	CVaR (α = 0.95)	Weight $\leq$ 1, No shorting	NSGA-II + GA	MSE, CVaR reduction
Mba & Angaman (2023)[21]	Cryptocurrencies	2020–2022	VaR (α = 0.90)	Market cap bounds	Entropy Pooling + K-Means	Diversification Index
Yu & Liu (2021)[15]	Stocks, Bonds, Mutual Funds	2010–2019	Mean-CVaR ($\alpha = 0.95$)	Risk tolerance bounds	PSO + Copula-GARCH	Sharpe Ratio, CVaR
Bulani et al. (2025)[22]	Stocks & Crypto	2015–2023	–	Long-only	PSO	Portfolio Stability
Kaut (2021)[23]	Historical Market Data	2000–2019	CVaR (α = 0.95)	Moment matching	MIP + K-Means	Wasserstein Distance
Jain et al. (2025)[17]	Nifty 100 (ESG)	2018–2024	CVaR (α = 0.95)	ESG constraint	TLBO + REM + TODIM	Sharpe, Gini–Simpson

Table 7 summarizes six key studies that collectively illustrate the methodological diversity in Mean–CVaR based portfolio optimization and clustering-driven asset allocation. Studies such as Bedoui et al. [16] and Yu and Liu [15] emphasize tail-risk minimization through CVaR or Mean–CVaR frameworks under high confidence levels ($\alpha = 0.95$), applying stochastic or evolutionary solvers like NSGA-II, GA, and PSO under bounded portfolio constraints to achieve robust optimization. In contrast, Mba and Angaman [21], Bulani et al. [22], and Kaut [23] utilize K-Means for asset grouping or scenario generation, enhancing diversification but often without explicit CVaR integration. Meanwhile, Jain et al. [17] introduces sustainability-driven constraints via ESG-based CVaR modeling, reflecting a conceptual shift toward responsible portfolio design. Overall, the studies demonstrate complementary strengths Mean–CVaR excels in quantifying downside risk, while K-Means enhances structural diversification but their integration within a single optimization framework remains limited, revealing a significant methodological gap for future exploration.

The comparative patterns identified in Table 6 reveal an ongoing methodological convergence between quantitative risk modeling and data-driven asset structuring. While existing works differ in datasets, algorithms, and constraints, they share a unified objective of balancing risk minimization with portfolio diversification. This synergy suggests that integrating Mean–CVaR's risk sensitivity with K-Means' clustering adaptability could provide a more comprehensive framework for dynamic asset allocation. The following section, Synthesis of Findings, further elaborates on this relationship, organizing the reviewed studies into conceptual categories and outlining methodological pathways for developing hybrid risk clustering portfolio models.

5) Synthesis of Findings

The synthesis of the six reviewed studies reveals three main methodological orientations in portfolio optimization research integrating Mean–CVaR and clustering techniques. The first group, represented by Bedoui et al. [16] and Yu and Liu [15], emphasizes tail-risk optimization, applying Mean-CVaR with stochastic or evolutionary solvers such as NSGA-II, PSO, and Copula-GARCH to minimize downside deviation under specific portfolio constraints. While mathematically rigorous, these models remain computationally intensive and lack structural diversification mechanisms.

The second orientation involves clustering-based diversification, as seen in Mba and Angaman [21], Bulani et al. [22], and Kaut [23], where K-Means clustering is used to identify asset similarity before optimization. This approach effectively enhances portfolio balance and reduces redundancy, yet clustering typically acts as a preprocessing stage rather than being embedded in the optimization model, limiting its integration with CVaR-based risk functions.

The third group, exemplified by Jain et al. [17], introduces hybrid learning-driven frameworks, combining Mean–CVaR optimization with sustainability and AI-based extensions such as TLBO, REM, and ESG constraints. This trend signifies a transition toward multi-objective and adaptive portfolio modeling that aligns risk management with environmental and ethical considerations.

Overall, the reviewed literature can be classified into a taxonomy of integration pathways: (1) Pre-clustering optimization K-Means applied before Mean–CVaR to enhance diversification, (2) Hybrid optimization iterative or joint integration of both methods, and (3) Scenario-based modeling clustering applied to data-driven scenario generation. This taxonomy highlights that existing studies only partially link risk quantification and structural diversification, leaving room for a unified hybrid framework that mathematically integrates Mean-CVaR and K-Means within a single optimization process.

4 Discussion

Based on the comparative analysis of the six eligible studies, several methodological and conceptual gaps are identified. First, although Conditional Value-at-Risk (CVaR) is consistently applied as a dominant tail-risk measure in portfolio optimization, direct integration of the Mean-CVaR model with clustering algorithms such as K-Means within a unified optimization framework remains scarce. Existing studies tend to employ CVaR alongside stochastic programming, metaheuristics, or copula-based simulations, yet clustering is mainly used for preprocessing or scenario generation rather than for dynamic portfolio weighting. This fragmentation highlights a methodological disconnect between risk quantification and asset grouping. Second, the empirical scope of current studies remains narrow, focusing primarily on single-asset portfolios (e.g., equities or cryptocurrencies). The limited exploration of heterogeneous multi-asset settings where volatilities, correlations, and downside risks differ substantially restricts the generalizability of existing models. Future studies should validate Mean-CVaR and K-Means hybrids across multiple asset classes to test cross-asset stability and diversification efficiency. Third, although various optimization solvers have been utilized including NSGA-II, PSO, TLBO, and MIP their simultaneous integration with clustering-based scenario generation is minimal. A hybridized pipeline that couples scenario generation (via K-Means or entropy-based grouping) with Mean-CVaR optimization could enable adaptive risk control and enhance robustness under regime shifts. Fourth, the inclusion of non-financial factors, particularly ESG (Environmental, Social, and Governance) criteria, is emerging but underdeveloped. Among the reviewed works, only Jain et al. [17] incorporated ESG constraints, yet without integrating them into a risk-return optimization under Mean-CVaR. The introduction of ESG-adjusted CVaR or multi-objective Mean-CVaR–ESG models represents a promising direction, especially for sustainable and responsible investment portfolios.

Finally, given that only six integration-relevant studies were identified, the synthesis here emphasizes depth over breadth. These papers collectively cover diverse mathematical formulations, asset classes, and optimization algorithms (as detailed in Table 7), but the evidence base remains modest. Hence, generalizations should be interpreted cautiously, while emphasizing the conceptual convergence between tail-risk minimization and data-driven clustering as a foundation for future empirical testing.

Based on these identified gaps, future research is recommended to:

a) Develop hybrid frameworks that explicitly combine Mean-CVaR optimization with K-Means based asset clustering or scenario generation, enabling simultaneous risk estimation and portfolio grouping.
b) Extend model validation to multi-asset datasets encompassing equities, bonds, commodities, and cryptocurrencies with heterogeneous risk dynamics. Future empirical validation should also incorporate datasets from emerging markets such as the ASEAN region to examine cross-market robustness and regional diversification effects.
c) Incorporate advanced metaheuristic or machine-learning-based solvers (e.g., PSO–K-Means, NSGA-II–CVaR hybrids) to address high-dimensional and non-convex optimization problems.
d) Conduct robustness testing across multiple market regimes (bullish, bearish, and crisis conditions) to evaluate model stability and resilience of optimized weights.
e) Integrate ESG-adjusted CVaR functions to align financial risk assessment with sustainability objectives, enhancing the real-world applicability and long-term relevance of portfolio optimization.

In summary, the synthesis underscores that integrating Mean-CVaR with clustering and sustainability-oriented extensions offers a fertile ground for methodological advancement. This convergence promises not only improved diversification efficiency and downside protection but also alignment with the evolving paradigm of sustainable, data-driven financial decision-making.

5 Conclusion

This study conducted a Systematic Literature Review (SLR) to map the evolution and methodological integration of the Mean Conditional Value-at-Risk (Mean-CVaR) framework and K-Means Clustering in financial asset portfolio optimization. Using six rigorously selected studies published between 2021 and 2025, supported by bibliometric analysis from Scopus, ScienceDirect, and Dimensions, this review reveals that CVaR remains the most dominant and coherent risk measure for capturing tail risk in portfolio optimization. However, its direct integration with K-Means Clustering within a unified optimization model is still limited. Most existing studies employ K-Means merely as a pre-clustering or scenario-generation tool, while the actual weight optimization is handled by metaheuristic or stochastic programming methods such as GA, PSO, TLBO, and MIP. Moreover, applications to multi-asset portfolios and the incorporation of non-financial dimensions such as Environmental, Social, and Governance (ESG) factors remain underexplored.

The synthesis highlights a clear research gap in bridging these two methodological domains. Future research should advance toward hybrid optimization frameworks that combine Mean-CVaR risk modeling with K-Means-based clustering or scenario generation to achieve simultaneous asset grouping and risk minimization. Empirical validation should extend to multi-asset datasets including equities, bonds, commodities, and cryptocurrencies and incorporate data from emerging markets such as the ASEAN region to test cross-market robustness and diversification. Additionally, the adoption of metaheuristic or machine-learning-based solvers (e.g., PSO–K-Means, NSGA-II–CVaR hybrids) and ESG-adjusted CVaR formulations is recommended to enhance both computational efficiency and sustainability relevance. In summary, this study contributes a structured foundation for developing more adaptive and resilient portfolio optimization models. By integrating Mean-CVaR and K-Means within a hybrid, multi-asset, and sustainability-oriented framework, future research can better address the dual challenges of risk concentration and diversification efficiency in increasingly volatile and interconnected financial markets.

CRediT Authorship Contribution Statement

Alim Jaizul Wahid: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Writing–Original Draft Preparation. **Riaman:** Writing–Review & Editing, Visualization, Investigation, Project Administration, Funding Acquisition. **Sukono:** Validation, Supervision.

Declaration of Generative AI and AI-assisted technologies

The authors affirm that generative AI technology was utilized to support the preparation of this manuscript. Specifically, OpenAI's ChatGPT (version GPT-5) was employed to assist in tasks such as content refinement, language structuring, and improving clarity in the presentation of results. All outputs generated by the tool were critically reviewed, revised, and integrated by the authors to ensure accuracy, coherence, and alignment with academic standards. The final manuscript reflects the authors' independent intellectual contribution and professional judgment.

Declaration of Competing Interest

The authors declare no competing interests

Funding and Acknowledgments

This research is supported by Master Thesis Research Grant from Indonesian Ministry of Higher Education, Science, and Technology for the year of 2025 under contract number 1646/UN6.3.1/PT.00/2025.

Data and Code Availability

The bibliometric data supporting the findings of this study were obtained from licensed academic databases, namely Scopus, ScienceDirect, and Dimensions. Due to licensing restrictions, the raw datasets cannot be shared publicly. However, the processed data used for analysis and the R scripts (Bibliometrix and Biblioshiny procedures) are available from the corresponding author upon reasonable request, subject to confidentiality and database access agreements.

References

[1] J. Fadika and Y. A. Indra, "Peran pasar modal dalam meningkatkan minat investasi pada generasi muda di era digital," *Journal of Management and Innovation Entrepreneurship*, vol. 2, no. 1, pp. 1700–1712, 2024. DOI: 10.70248/jmie.v2i1.1430.

[2] Y. Yang, "Clustering-based financial big data analysis for portfolio optimization," in *Proceedings of the 2025 International Conference on Big Data*, Association for Computing Machinery, 2025, pp. 36–42. DOI: 10.1145/3727505.3727511.

[3] M. A. Riswana and G. W. Yasa, "Influence of investment knowledge, minimum capital, influencers, technological progress on students interest in investing in the capital market," *International Journal of Business, Economics & Management*, vol. 7, no. 1, pp. 18–26, 2024. DOI: 10.21744/ijbem.v7n1.2249.

[4] L. H. Nguyen, L. X. D. Nguyen, and E. Adegbite, "Does mean-cvar outperform mean-variance? theoretical and practical perspectives," *Theoretical and Practical Perspectives*, pp. 1–52, 2018. DOI: https://dx.doi.org/10.2139/ssrn.3143827.

[5] L. Gubu, E. Cahyono, H. Budiman, and M. K. Djafar, "Cluster analysis for mean-variance portfolio selection: A comparison between k-means and k-medoids clustering," *Jurnal Riset dan Aplikasi Matematika*, vol. 7, no. 2, pp. 104–115, 2023. DOI: 10.26740/jram.v7n2.p104-115.

[6] U. Arif, D. Muhammad, T. Sohail, and M. I. Majeed, "Portfolio optimization with mean-variance & mean-cvar: Evidence from pakistan stock market," *International Journal of Management Research and Emerging Sciences*, vol. 10, no. 2, pp. 215–226, 2020.

[7] S. Petchrompo, A. Wannakrairot, and A. K. Parlikad, "Pruning pareto optimal solutions for multi-objective portfolio asset management," *European Journal of Operational Research*, vol. 297, no. 1, pp. 203–220, 2022. DOI: 10.1016/j.ejor.2021.04.053.

[8] I. Lisnawati and R. Subekti, "Conditional value at risk (cvar) of portfolio using conditional copula," *Jurnal Kajian dan Terapan Matematika*, vol. 7, no. 3, pp. 64–72, 2018. DOI: 10.14710/j.gauss.v9i3.28913.

[9] M. Salahi, F. Mehrdoust, and F. Piri, "Cvar robust mean-cvar portfolio optimization," *ISRN Applied Mathematics*, vol. 2013, pp. 1–9, 2013. DOI: 10.1155/2013/570950.

[10] F. Soleymani and M. Vasighi, "Efficient portfolio construction by means of cvar and k-means++ clustering analysis: Evidence from the nyse," *International Journal of Finance and Economics*, vol. 27, no. 3, pp. 3679–3693, 2022. DOI: 10.1002/ijfe.2344.

[11] M. Gulliksson, S. Mazur, and A. Oleynik, "Minimum var and minimum cvar optimal portfolios: The case of singular covariance matrix," *Results in Applied Mathematics*, vol. 26, p. 100 557, 2025. DOI: 10.1016/j.rinam.2025.100557.

[12] N. Abudurexiti, K. He, D. Hu, S. T. Rachev, H. Sayit, and R. Sun, "Portfolio analysis with mean-cvar and mean-cvar-skewness criteria based on mean–variance mixture models," *Annals of Operations Research*, vol. 336, no. 1-2, pp. 945–966, 2024. DOI: 10.1007/s10479-023-05396-1.

[13] M. M. Kowsar, M. Mohiuddin, and S. Islam, "Mathematics for finance: A review of quantitative methods in loan portfolio optimization," *International Journal of Scientific Interdisciplinary Research*, vol. 4, no. 3, pp. 1–29, 3. Available online.

[14] X. Guo, R. H. Chan, W. K. Wong, and L. Zhu, "Mean–variance, mean–var, and mean–cvar models for portfolio selection with background risk," *Risk Management*, vol. 21, no. 2, pp. 73–98, 2019. DOI: 10.1057/s41283-018-0043-2.

[15] C. Yu and Y. Liu, "A personalized mean-cvar portfolio optimization model for individual investment," *Mathematical Problems in Engineering*, vol. 2021, p. 12, 2021. DOI: 10.1155/2021/8863597.

[16] R. Bedoui, R. Benkraiem, K. Guesmi, and I. Kedidi, "Portfolio optimization through hybrid deep learning and genetic algorithms vine copula-garch-evt-cvar model," *Technological Forecasting and Social Change*, vol. 197, 2023. DOI: 10.1016/j.techfore.2023.122887.

[17] V. Jain, R. R. Sahay, and Nupur, "Integrated robust em and todim approach with teaching learning based portfolio optimization," *Computational Economics*, 2025. DOI: 10.1007/s10614-025-11013-z.

[18] H. Firmansyah and D. Rosadi, "Construction of stock portfolios based on k-means clustering of continuous trend features," *Jurnal Matematika Integratif*, vol. 20, no. 2, pp. 149–172, 2024. DOI: 10.24198/jmi.v20.n2.53351.149-172.

[19] R. Siagian, P. Sirait, and A. Halima, "E-commerce customer segmentation using k-means algorithm and length, recency, frequency, monetary model," *Journal of Informatics and Telecommunication Engineering*, vol. 5, no. 1, pp. 21–30, 2021. DOI: 10.31289/jite.v5i1.5182.

[20] D. Wu, X. Wang, and S. Wu, "Construction of stock portfolios based on k-means clustering of continuous trend features," *Knowledge-Based Systems*, vol. 252, no. 2022, p. 109 358, 2022. DOI: 10.1016/j.knosys.2022.109358.

[21] J. C. Mba and E. S. E. F. Angaman, "A k-means classification and entropy pooling portfolio strategy for small and large capitalization cryptocurrencies," *Entropy*, vol. 25, no. 8, 2023. DOI: 10.3390/e25081208.

[22] V. Bulani, M. Bezbradica, and M. Crane, "Improving portfolio management using clustering and particle swarm optimisation," *Mathematics*, vol. 13, no. 10, 2025. DOI: 10.3390/math13101623.

[23] M. Kaut, "Scenario generation by selection from historical data," *Computational Management Science*, vol. 18, no. 3, pp. 411–429, 2021. DOI: 10.1007/s10287-021-00399-4.

[24] M. J. Page et al., "The prisma 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, n71, 2021. DOI: 10.1136/bmj.n71.

[25] A. Z. Irmansyah, D. Chaerani, and E. Rusyaman, "A systematic review on integer multi-objective adjustable robust counterpart optimization model using benders decomposition," *JTAM (Jurnal Teori dan Aplikasi Matematika)*, vol. 6, no. 3, pp. 678–690, 2022. DOI: 10.31764/jtam.v6i3.8578.