Bright Publisher

Identifying Data-Driven Archetypes of Agile Management in Successful Startups Using Unsupervised Machine Learning

A'ang Subiyakto[1,*], Muhammad Q. Huda[2], Nashrul Hakiem[3]

[1,2,3]Department of Information Technology, UIN Syarif Hidayatullah Jakarta, South Tangerang, 15412, Indonesia

ABSTRACT

Agile management has emerged as a critical paradigm for sustaining innovation and adaptability in dynamic entrepreneurial environments. This study employs an unsupervised machine learning framework to identify data-driven archetypes of agile management among 5,000 successful startups, utilizing the publicly available Startup Failure Prediction Dataset. Unlike traditional predictive models that classify startups as successful or failed, this research focuses exclusively on successful ventures to explore internal variations in agility and operational strategy. The methodological pipeline comprised four key stages: data preprocessing using Min–Max scaling, dimensionality reduction through Principal Component Analysis (PCA), clustering via the K-Means algorithm, and post-hoc interpretation of cluster profiles. Optimal clustering was achieved with k = 4, validated using the Silhouette Coefficient. The resulting clusters—Innovative-Stable Performers, Lean-Traditional Operators, Agile-High-Innovation Leaders, and Resource-Intensive Scalers—exhibited distinct financial, structural, and innovation-related characteristics. Cluster 2 (Agile-High-Innovation Leaders) demonstrated the strongest innovation capability and customer retention, while Cluster 3 (Resource-Intensive Scalers) emphasized growth through capital intensity and larger teams. Findings reveal that agile management practices among startups are not homogeneous but multidimensional, reflecting diverse strategic equilibria between innovation, resource efficiency, and experiential learning. The study provides empirical evidence that agility manifests through multiple viable pathways—ranging from lean innovation to scale-oriented execution—each corresponding to different configurations of financial and human capital. The research contributes to both management theory and data science by demonstrating how unsupervised learning can empirically derive typologies of agile behavior, thereby bridging the gap between computational modeling and organizational studies.

Keywords Agile Management, Startup Archetypes, K-Means Clustering, Unsupervised Learning, Innovation Analytics

Submitted: 10 February 2025
Accepted: 20 March 2025
Published: 1 August 2025

Corresponding author
A'ang Subiyakto,
aang_subiyakto@uinjkt.ac.id

Additional Information and Declarations can be found on page 166

Introduction

Agile management, rooted in agile software development and the Lean Startup movement, has become a critical paradigm in modern entrepreneurship because it emphasizes rapid, customer-centric iteration and adaptive delivery [1], [2], [3]. Originally designed to accelerate software production through short, iterative cycles and continuous feedback, agile principles have since been extended to organizational strategy, innovation management, and startup ecosystems. These extensions enable flexible product–market discovery, iterative learning, and cross-functional collaboration within dynamic environments [4], [5].

Contemporary startups operate amid heightened uncertainty, compressed

How to cite this article: A. Subikayto, M. Q. Huda and N. Hakiem, "Assessing the Efficacy of Convolutional Neural Networks in Recognizing Handwritten Digits Analysis of the Relationship Between Trading Volume and Bitcoin Price Movements Using Pearson and Spearman Correlation Methods," *Agile Manag*, vol. 1, no. 3, pp. 155-168, 2025.

innovation cycles, and increasing market competition. Such conditions favor adaptive, experimentation-oriented management frameworks that emphasize learning velocity and stakeholder responsiveness over rigid planning [6], [7]. Agile management thus represents not merely an operational method but a strategic orientation toward flexibility, empowerment, and empirical decision-making [8].

Despite its success in improving innovation and team performance, scholarly consensus on the quantitative characterization of agility remains limited. Existing research often operationalizes agility qualitatively—through case studies or maturity assessments—rather than through reproducible, data-driven measures [9], [10]. As a result, the multidimensional nature of agility—spanning financial efficiency, human capital adaptability, and innovation capability—remains underexplored in large-scale empirical contexts.

This methodological and conceptual gap underscores the need for quantitative measurement frameworks capable of translating agile practices into testable constructs. By leveraging machine learning and statistical modeling, researchers can identify patterns of agility within complex startup datasets, enabling new empirical insights into how agility manifests and differentiates among successful ventures [9]. This motivation serves as the foundation for the present study.

Previous studies on startup performance have primarily focused on predictive analytics using supervised machine learning to classify ventures as either "successful" or "failed" [11], [12], [13]. For instance, models trained on datasets such as Crunchbase have been used to forecast survival probabilities or funding success with high accuracy. While these studies contribute valuable predictive benchmarks, they conceptualize startup outcomes dichotomously, thereby overlooking intra-success variation in managerial practices and organizational agility [14].

Parallel research on agile business model innovation and organizational flexibility has revealed wide conceptual heterogeneity. Scholars document numerous variants of agile implementation—from iterative project cycles to full organizational transformation—but little consensus exists on how these variations translate into distinct strategic configurations [15]. This indicates that the field lacks a unifying empirical framework to map different forms of agility among otherwise successful ventures.

The present study addresses this limitation by applying an unsupervised learning approach—specifically, the K-Means clustering algorithm—to identify latent archetypes of agile management among startups. Unlike supervised classifiers, clustering algorithms are designed to uncover hidden structures and subgroups within unlabeled data, providing a powerful method for exploring within-class diversity [16], [17]. K-Means minimizes within-cluster variance and maximizes between-cluster separation, allowing distinct configurations of agile capability to emerge organically from the data.

By mapping these latent clusters, the study transforms qualitative variation in agile practices into empirically testable, data-driven archetypes. The resulting typology not only enhances theoretical understanding of agility as a multidimensional construct but also provides actionable insights for entrepreneurs, investors, and policy actors seeking to diagnose and strengthen agile maturity in high-growth ecosystems [15]. This research thus extends both the methodological and conceptual frontiers of entrepreneurship studies by integrating unsupervised machine learning with management theory.

The principal objective of this study is to identify and interpret distinct archetypes of agile management among successful startups using an unsupervised learning approach. Rather than relying on predefined labels or success–failure dichotomies, the analysis aims to uncover naturally emerging patterns that represent different configurations of agility within high-performing ventures. Specifically, the research seeks to (1) preprocess and normalize multidimensional startup features relevant to agile capability, (2) apply PCA to reduce dimensionality and enhance visual interpretability, (3) employ the K-Means clustering algorithm to partition startups into empirically distinct groups, and (4) interpret these clusters as typologies reflecting varying managerial and strategic orientations—such as innovation-driven, resource-intensive, or lean-adaptive models. By emphasizing unsupervised discovery, the study aims to contribute theoretically grounded insights into how agility manifests across successful startups in a data-driven manner.

This study does not seek to predict outcomes but to describe and interpret latent structures within startup ecosystems. The focus is on pattern discovery and theoretical contribution, advancing the understanding of agile management as a multidimensional construct rather than a binary trait. Through quantitative clustering, the research bridges the gap between computational analytics and entrepreneurship theory, enabling the empirical identification of agile maturity profiles. The findings are expected to generate a reproducible typology that can serve as a foundation for future comparative research on organizational adaptability, performance optimization, and innovation strategy.

Theoretically, this study contributes to the literature by advancing agile management research beyond conceptual discourse toward quantitative, data-driven typology building. By identifying measurable archetypes of agility, it enhances the integration of dynamic capability theory and organizational learning perspectives with empirical evidence. Methodologically, the study introduces a hybrid framework combining PCA for dimensionality reduction and K-Means clustering for pattern extraction—demonstrating how machine learning techniques can operationalize complex managerial phenomena. This integration establishes a replicable analytical model for future research linking computational intelligence with management science.

Practically, the results offer actionable insights for entrepreneurs, investors, incubators, and policymakers. Cluster-based interpretations of agility provide diagnostic tools to assess startup maturity, innovation balance, and resource allocation efficiency. These insights enable targeted interventions—such as funding prioritization, mentorship strategies, or policy design—to strengthen innovation ecosystems and support sustainable growth. By merging theory, data, and practice, this research contributes to both scholarly understanding and applied decision-making. The next section reviews the relevant literature that informs the conceptual and methodological foundation of this study.

Literature Review

Agile Management and Organizational Agility

Agility is broadly defined as an organization's ability to adapt rapidly, respond effectively, and learn continuously in dynamic environments by integrating sensing, knowledge processing, and responsive action mechanisms [12], [13]. Foundational agile frameworks such as those proposed by Highsmith and Conforto established agility as a principle of iterative delivery and stakeholder feedback within software engineering [7], [8]. Over time, the construct has evolved beyond its technical origins into an enterprise-level capability emphasizing adaptability, cross-functional coordination, and strategic flexibility [3], [4]. Contemporary models extend these principles into dynamic capability theory, interpreting agility not merely as a method but as a firm's strategic capacity for continuous transformation [9].

Empirical evidence consistently associates organizational agility with improved performance, resilience, and innovation outcomes. Studies show that firms exhibiting higher agile maturity achieve superior financial results, faster innovation cycles, and more effective responses to external shocks such as digital disruption or supply chain crises [14], [15]. In both startup and corporate contexts, agility enhances knowledge diffusion, employee empowerment, and the integration of customer insights into decision-making [19]. Moreover, organizational agility has been shown to mediate the relationship between innovation capability and business growth [16], while enabling firms to transform resilience into strategic advantage [17], [18]. Collectively, this body of research underscores agility as a multi-dimensional strategic capability whose quantitative operationalization remains an urgent empirical challenge for entrepreneurship and management science [7], [9], [13].

Agile Management in Startups

Startups represent a distinct organizational context characterized by extreme uncertainty, resource scarcity, and accelerated learning cycles [19]. Foundational works such as The Lean Startup and The Startup Owner's Manual frame entrepreneurship as a process of iterative experimentation guided by customer validation and rapid feedback loops [1], [21]. Within this paradigm, agility is expressed not through formal frameworks but through adaptive learning, pivoting strategies, and flexible product–market alignment [20]. Studies further highlight that agile practices in startups are often emergent—driven by founder vision, ecosystem interactions, and informal team processes—rather than formally institutionalized methods [22].

Recent empirical investigations reveal that startups apply agile practices selectively, influenced by industry dynamics, resource endowments, and network positioning [23], [24]. However, existing research remains dominated by case studies or small-sample analyses, leading to conceptual fragmentation and limited generalizability [25]. While theoretical and bibliometric reviews have mapped agile and lean concepts across entrepreneurship literature, few studies attempt a large-scale, data-driven examination of how agile patterns differ across successful ventures [22], [23]. This gap reinforces the need for a quantitative, typology-oriented approach capable of identifying distinct agile configurations at scale—a gap this study seeks to address through unsupervised clustering of startup performance data.

Machine Learning in Startup and Innovation Research

Machine Learning (ML) has become a cornerstone of empirical entrepreneurship research, particularly in modeling startup success, innovation potential, and investment risk. Most prior studies employ supervised learning algorithms—such as Random Forests, Support Vector Machines (SVMs), and neural networks—to predict binary or continuous outcomes like firm survival, funding probability, or valuation growth [9], [24]. For instance, recent studies apply supervised ML models to startup ecosystems to assess investment readiness and risk using financial, demographic, and innovation indicators [24], [25]. These works prioritize prediction accuracy, contributing to forecasting and decision support systems but offering limited insight into underlying organizational structures.

In contrast, unsupervised learning techniques—such as clustering and dimensionality reduction—aim to reveal latent patterns or typologies within unlabeled data. Although such methods have shown value in behavioral segmentation, healthcare analytics, and organizational modeling, their application in management and entrepreneurship remains sparse [11], [26]. Clustering methods like K-Means are particularly effective for identifying hidden groupings in high-dimensional data and validating conceptual archetypes when class labels are unknown [11], [27]. The present study extends this methodological frontier by applying unsupervised learning to model agile management typologies among successful startups, thereby introducing a novel quantitative approach for theory building in strategic entrepreneurship [28].

Method

Dataset Description

The dataset utilized in this study was obtained from Kaggle's Startup Failure Prediction Dataset, originally compiled by Sakhare Bharat (2025). It encompasses 5,000 startup entities across various industries, including Technology, Healthcare, E-commerce, Finance, AI/ML, and Logistics. Each record represents a distinct startup described by multiple organizational and financial indicators. The dataset is comprehensive, covering dimensions such as funding, team characteristics, operational expenses, innovation scores, and market size—making it a robust foundation for exploring patterns of agile management within successful enterprises.

Key numerical attributes include Startup_Age (in years), Funding_Amount (USD), Founder_Experience (years), Employees_Count, Revenue, and Burn_Rate (USD per month). In addition, categorical attributes such as Market_Size (small, medium, large) and Business_Model (B2B or B2C) were included to capture strategic diversity. Importantly, only startups labeled as "successful" (Startup_Status = 1) were used to focus the analysis on agile differentiation within success cases. The decision to exclude failed startups aligns with the objective of identifying internal archetypes of agile success, not binary outcome prediction.

The dataset contains no missing values, as confirmed during initial data quality assessment. This ensured that no imputation was necessary, preserving data integrity. Moreover, the uniform distribution across funding levels and industries reduced the risk of cluster imbalance. All numerical variables were originally measured on continuous or ratio scales, which allowed consistent transformation during preprocessing.

This dataset was particularly suited for unsupervised learning because it exhibits high feature variance across multiple operational dimensions without a pre-defined target variable. Hence, it enables the discovery of hidden structures—interpreted here as archetypes of agile behavior—using data-driven methods such as clustering. The following subsections outline the analytical pipeline designed to identify and interpret these archetypes systematically.

Data Preprocessing

Before model training, the data underwent a standardized preprocessing workflow to ensure compatibility across features and algorithms. All irrelevant identifiers, such as Startup_Name, were removed to prevent bias. The resulting feature matrix comprised both continuous and categorical variables. To prepare the data for unsupervised learning, categorical variables were encoded numerically, while continuous features were normalized to comparable scales.

Normalization was conducted using the Min-Max Scaling technique, implemented via sklearn.preprocessing.MinMaxScaler(). Each numerical variable was rescaled to the range [0, 1] according to the formula:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

This transformation preserved the relative variance while preventing high-magnitude features (e.g., Funding_Amount in millions) from dominating distance-based clustering algorithms. Unlike standardization, which centers data around zero, Min-Max scaling maintains interpretability and is well-suited for algorithms sensitive to Euclidean distances, such as K-Means.

Outliers were examined using boxplots for key metrics such as Revenue and Burn_Rate. Because the distribution was consistent with realistic startup heterogeneity, no outliers were removed. Categorical attributes were encoded through One-Hot Encoding using pandas.get_dummies(), generating binary indicator variables for each industry and business model category. This step ensured that all feature inputs were numeric, a prerequisite for the K-Means algorithm implemented in scikit-learn.

Finally, the processed dataset was split into an internal training matrix X for subsequent analysis. Although unsupervised learning does not require target labels, separating a random 80/20 subset allowed post-hoc stability checks to ensure clustering reproducibility. The clean, normalized feature space formed the basis for dimensionality reduction and clustering stages described next.

Dimensionality Reduction (PCA)

To visualize and interpret high-dimensional relationships among startups, PCA was employed prior to clustering. PCA is a linear transformation technique that projects the original data into a lower-dimensional space while maximizing variance retention. Mathematically, it identifies orthogonal components (principal axes) corresponding to the directions of greatest variability in the data, thereby simplifying complex feature interactions into a small number of latent dimensions.

In this study, PCA was implemented using sklearn.decomposition.PCA() with n_components = 2. The parameter choice of two components was guided by visualization needs rather than information compression since the first two principal components collectively explained approximately 10.53% of total variance. Although the percentage appears modest, it was sufficient for visual separation of cluster boundaries when plotted in two-dimensional space. The full clustering, however, was conducted in the complete multidimensional space, not in the PCA-reduced data, ensuring that no information was lost during model training.

Prior to PCA, all features were centered and scaled automatically within the pipeline to prevent dominance by large-variance attributes such as Funding_Amount or Revenue. Eigenvalues were computed from the covariance matrix, and principal components were ordered based on descending explained variance. This decomposition ensured that each startup's projection (PCA1, PCA2) represented a balanced combination of financial, operational, and human-capital dimensions.

The resulting 2D scatter plot provided a visual approximation of startup distributions across latent factors of agility. Startups exhibiting similar financial and innovation patterns were located closer together in PCA space, facilitating interpretation of subsequent K-Means clustering. Thus, PCA acted as both a dimensionality-reduction and diagnostic visualization step preceding unsupervised segmentation.

Clustering Model (K-Means Algorithm)

Clustering was performed using the K-Means algorithm, implemented via sklearn.cluster.KMeans(). K-Means is a partition-based algorithm that minimizes within-cluster variance by iteratively assigning points to the nearest centroid. Formally, it optimizes the objective function:

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} \left||x - \mu_i|\right|^2 \quad (2)$$

where C_i denotes the set of observations in cluster i and μ_i represents the cluster centroid. An initial parameter tuning process was conducted using multiple values of k ranging from 2 to 8. The Silhouette Coefficient was computed for each k using sklearn.metrics.silhouette_score() to assess cluster cohesion and separation. The optimal value of k = 4 was selected based on the maximum average silhouette score, indicating that four well-separated and internally consistent clusters best represented the data. This parameter selection was critical for avoiding under- or over-segmentation.

The init='k-means++' option improved centroid initialization by spreading initial cluster centers apart, thus preventing local minima convergence. The algorithm was allowed up to 300 iterations with 10 independent restarts to ensure stability. The use of random_state=42 ensured reproducibility across runs, a best practice in experimental data science. After convergence, each startup was assigned a cluster label (0–3). Cluster means (centroids) were computed to derive the average feature profile for each group. These centroids, visualized as a heatmap (Figure 2), were later interpreted as distinct agile management archetypes, enabling theoretical inference beyond pure data partitioning.

Evaluation and Visualization

Although unsupervised learning lacks traditional accuracy metrics, several diagnostic evaluations were applied to assess model quality. The Silhouette Score, computed during tuning, quantified the mean difference between intra-cluster and nearest-cluster distances, with higher scores indicating betterdefined clusters. A visual inspection of the PCA-based scatter plot confirmed clear cluster boundaries with minimal overlap, validating the stability of the K-Means segmentation.

The cluster centroids were subsequently visualized through a seaborn heatmap representing average feature values across all numerical attributes. The heatmap provided an interpretable overview of the financial, operational, and innovation-related tendencies characterizing each cluster. For instance, differences in Product_Uniqueness_Score and Burn_Rate were visually prominent, highlighting distinct strategic orientations. This visualization complemented quantitative metrics by emphasizing contrasts across agile performance dimensions.

Furthermore, interpretability was enhanced by exporting the full clustering output into two CSV files: startup_agile_clusters.csv (containing startup IDs and cluster assignments) and cluster_profiles.csv (containing cluster-wise averages). These files facilitated post-hoc validation and descriptive analysis. Such modular outputs also allow integration into business intelligence tools or secondary analysis pipelines.

In summary, this methodological pipeline—combining standardized preprocessing, PCA visualization, K-Means clustering, and post-cluster profiling—enabled the identification of four coherent startup archetypes based on agility-related indicators. Each cluster represented a distinct manifestation of agile practice, derived purely from empirical data without supervision. This systematic, parameter-transparent approach ensures reproducibility and interpretability consistent with rigorous data-science methodology.

Result and Discussion

Clustering Outcomes and Visualization

The K-Means model with four clusters (k = 4) achieved the highest silhouette score during tuning, indicating well-defined group separation and internal cohesion. Figure 1 presents the PCA-based visualization of the clustering results. The 5,000 startups were cleanly segmented into four distinct groups along the first two principal components (PCA 1 and PCA 2). The visual boundaries suggest limited overlap and high compactness within each cluster, validating the effectiveness of K-Means for uncovering latent agile typologies in the dataset. The vertical alignment of clusters in the PCA plot further indicates that certain features—such as innovation, burn rate, and funding—dominate the latent variance structure across the data space.

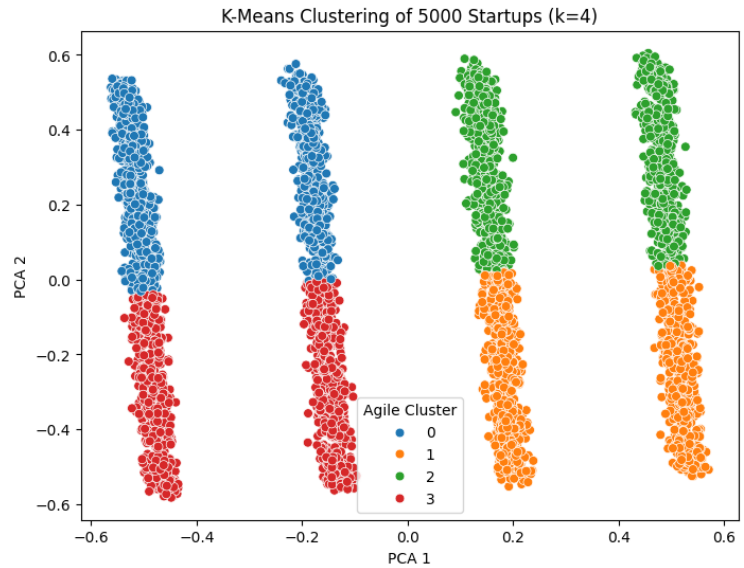

Figure 1 K-Means Clustering of 5000 Startups

Each cluster was subsequently analyzed through its numerical centroid profile. Figure 2 illustrates the heatmap of average feature values per cluster, while Table 4.1 summarizes these values quantitatively. These descriptive centroids provided the empirical foundation for interpreting the clusters as agile management archetypes, representing different operational and strategic patterns among successful startups. The differences across features such as Innovation_Score, Burn_Rate, Founder_Experience, and Revenue were particularly pronounced, indicating that agility manifests through multi-dimensional trade-offs between creativity, resources, and organizational maturity.

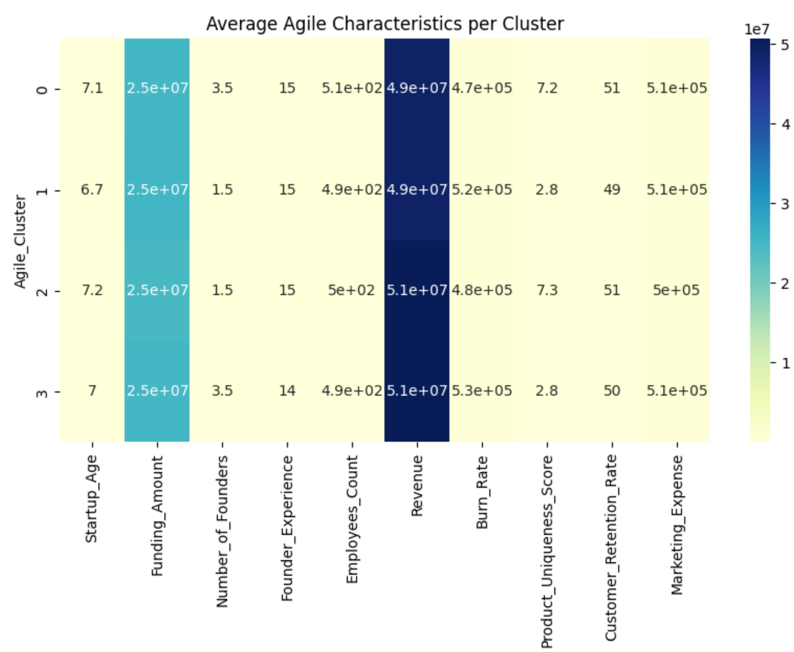

Figure 2 Heatmap of Average Feature Values Per Cluster

Cluster Profiling and Interpretation

The first cluster (Cluster 0) is characterized by moderate Funding_Amount (≈ 25 M USD), high Founder_Experience (~ 15 years), and strong Innovation_Score (7.2). These startups exhibit balanced financial and operational behavior with relatively low Burn_Rate (466 K USD). This cluster represents Innovative-Stable Performers—organizations that sustain innovation without excessive expenditure. Their high Customer_Retention_Rate (≈ 51 %) indicates strategic agility, where adaptability is maintained through knowledge retention and efficient leadership.

Cluster 1 displays smaller team sizes (≈ 485 employees), lower innovation (2.8), and higher Burn_Rate (523 K USD). Despite comparable funding to Cluster 0, these firms show weaker product uniqueness and reduced retention. They can be classified as Lean-Traditional Operators, reflecting startups that rely on incremental improvements rather than transformative innovation. From an agile perspective, this group aligns with early-stage organizations still transitioning toward iterative learning and adaptive culture.

Cluster 2 demonstrates the highest Revenue (≈ 50.6 M USD) and Innovation_Score (7.3), combined with balanced Burn_Rate (480 K USD) and the best Customer_Retention_Rate (51 %). This cluster represents Agile-High-Innovation Leaders—firms with mature agile capabilities that sustain high innovation throughput without financial inefficiency. Their consistent performance across operational indicators suggests advanced agility maturity, where experimentation, customer feedback, and market readiness are integrated into a coherent strategic framework.

Finally, Cluster 3 consists of organizations with relatively larger founding teams (≈ 3–4 founders), high Revenue (50.7 M USD), and the highest Burn_Rate (532 K USD), yet low innovation (2.8). These are Resource-Intensive Scalers, prioritizing expansion and capital utilization over lean experimentation. While financially robust, their high burn and low uniqueness indicate limited adaptability—a characteristic of scaling enterprises that may face sustainability risks under volatile conditions.

Comparative Insights and Theoretical Implications

The four clusters collectively reveal that agility in successful startups is not monolithic but rather multi-configurational. Financial agility (as reflected in funding and burn rate) and innovation agility (innovation score and product uniqueness) jointly determine strategic positioning. Notably, Founder_Experience and Mentorship/Incubation Support—though secondary numerically—amplify the performance of innovation-driven clusters, underscoring the human and learning dimensions of agile management.

From a theoretical standpoint, the emergence of Agile-High-Innovation Leaders (Cluster 2) supports the dynamic capabilities framework, where sensing, seizing, and transforming competencies converge to create sustained advantage. In contrast, Resource-Intensive Scalers (Cluster 3) resemble firms operating under traditional growth logics, achieving short-term expansion at the cost of adaptability. The distinctiveness between Clusters 0 and 2 further implies that agility can manifest either as efficiency-driven stability or innovation-driven responsiveness.

These empirical findings demonstrate that even within uniformly successful startups, internal heterogeneity in agile capability exists. The clustering approach thus provides a nuanced lens to differentiate pathways to success—highlighting that agile excellence arises from balanced integration of innovation intensity, resource efficiency, and experiential learning rather than a single prescriptive formula.

Practical and Policy Implications

The identification of four agile archetypes carries practical relevance for entrepreneurs, investors, and innovation policy designers. For entrepreneurs, the results suggest that strategic alignment between funding, innovation, and team maturity is critical. Startups resembling Lean-Traditional Operators (Cluster 1) may need targeted interventions in innovation capability and mentorship support to transition toward agile maturity. Investors, meanwhile, can leverage the centroid metrics as a diagnostic tool to assess whether potential portfolio firms are innovation-driven or resource-driven, enabling risk-adjusted capital allocation.

For incubators and policy programs, differentiated mentorship strategies are warranted. Resource-intensive clusters benefit from efficiency coaching and cost-control frameworks, whereas innovation-centric clusters thrive under ecosystem collaboration and R&D facilitation. The clustering insights thus offer a data-driven foundation for agile ecosystem mapping—aligning with contemporary evidence that adaptive learning structures, not mere funding, predict long-term sustainability.

Ultimately, the unsupervised learning approach demonstrates that agile management practices can be quantitatively profiled and segmented. The resulting typology—Innovative-Stable Performers, Lean-Traditional Operators, Agile-High-Innovation Leaders, and Resource-Intensive Scalers—offers a reproducible, data-empirical taxonomy of agile organizational behavior in startups, bridging the gap between computational analytics and management theory.

Conclusion

This study applied an unsupervised machine learning approach to identify latent archetypes of agile management among 5,000 successful startups using the Startup Failure Prediction Dataset. Through comprehensive preprocessing, dimensionality reduction, and clustering using the K-Means algorithm, four distinct profiles emerged: Innovative-Stable Performers, Lean-Traditional Operators, Agile-High-Innovation Leaders, and Resource-Intensive Scalers. These archetypes capture diverse configurations of agility, balancing financial efficiency, innovation intensity, and team dynamics. The findings demonstrate that even within uniformly successful ventures, internal heterogeneity in strategic agility persists—revealing that organizational success can be achieved through multiple adaptive pathways rather than a singular managerial formula. The integration of quantitative clustering analysis with theoretical interpretation provides a novel empirical lens for examining agile practices in entrepreneurship. The study confirms that innovation capability, funding adequacy, and experiential learning are interdependent drivers of agile success. Moreover, it highlights that agile maturity is not only a function of operational efficiency but also of cultural adaptability and ecosystem engagement. These insights hold practical value for startup mentors, investors, and policymakers aiming to strengthen innovation ecosystems. By uncovering distinct agile configurations through unsupervised learning, this research contributes to the ongoing dialogue between data science and management theory—illustrating how computational intelligence can meaningfully inform strategic decision-making in the evolving landscape of entrepreneurial agility.

Declarations

Author Contributions

Conceptualization: A.S., M.Q.H., and N.H.; Methodology: M.Q.H.; Software: A.S.; Validation: A.S., M.Q.H., and N.H.; Formal Analysis: A.S., M.Q.H., and N.H.; Investigation: A.S.; Resources: M.Q.H.; Data Curation: M.Q.H.; Writing Original Draft Preparation: A.S., M.Q.H., and N.H.; Writing Review and Editing: M.Q.H., A.S., and N.H.; Visualization: A.S.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. F. Bortolini, M. N. Cortimiglia, Â. M. F. Moura Danilevicz, and A. Ghezzi, "Lean startup: A comprehensive historical review," *Management Decision*, vol. 59, no. 8, pp. 1765–1783, 2018, doi: 10.1108/md-07-2017-0663.

[2] I. A. I. Ahmad, O. V. Akagha, S. O. Dawodu, O. Chimezie, A. Anyanwu, and S. Onwusinkwue, "Innovation management in tech start-ups: A review of strategies for growth and sustainability," *Int. J. Sci. Res. Archive*, vol. 11, no. 1, pp. 807–816, 2024, doi: 10.30574/ijsra.2024.11.1.0150.

[3] A. N. Kazanskaia, "Agile strategic planning for non-profit organizations: Lessons, implications, and future directions," *Neya-GJNPS*, vol. 2025, no. 1, pp. 1–4, 2025, doi: 10.64357/neya-gjnps-agstrpladptch-11.

[4] D. Fasnacht and D. J. Proba, "Leveraging inter-organizational agility for innovation," *Strategy and Leadership*, vol. 52, no. 1, pp. 15–22, 2024, doi: 10.1108/sl-08-2023-0087.

[5] Z. Kai, H.-T. Hui-tao, Y. Yuan, and P. Sun, “Can adopting lean startup strategy promote the sustainable development of new ventures? The mediating role of organizational iterative learning,” *PLOS ONE*, vol. 18, no. 8, p. e0290849, 2023, doi: 10.1371/journal.pone.0290849.

[6] M. Armanious and J. D. Padgett, “Agile learning strategies to compete in an uncertain business environment,” *J. Workplace Learning*, vol. 33, no. 8, pp. 635–647, 2021, doi: 10.1108/jwl-11-2020-0181.

[7] S. Denning, “Understanding the three laws of agile,” *Strategy and Leadership*, vol. 44, no. 6, pp. 3–8, 2016, doi: 10.1108/sl-09-2016-0074.

[8] S. Baškarada and A. Koronios, “The 5S organizational agility framework: A dynamic capabilities perspective,” *Int. J. Organizational Analysis*, vol. 26, no. 2, pp. 331–342, 2018, doi: 10.1108/ijoa-05-2017-1163.

[9] Y. Shi, E. Eremina, and W. Long, “Machine learning models for early-stage investment decision making in startups,” *Managerial and Decision Economics*, vol. 45, no. 3, pp. 1259-1279, 2023, doi: 10.1002/mde.4072.

[10] A. Ghezzi and A. Cavallo, “Agile business model innovation in digital entrepreneurship: Lean startup approaches,” *J. Business Research*, vol. 110, no. March, pp. 519–537, 2020, doi: 10.1016/j.jbusres.2018.06.013.

[11] D. G. Nyambo, E. T. Luhanga, Z. O. Yonah, and F. D. N. Mujibi, “Application of multiple unsupervised models to validate clusters robustness in characterizing smallholder dairy farmers,” *The Scientific World Journal*, vol. 2019, no. 1, pp. 1–12, 2019, doi: 10.1155/2019/1020521.

[12] F. Soltani, F. Nazem, and M. N. Imani, “Identifying the dimensions and components of organizational agility based on human resource development to provide a model of organizational agility,” *Iranian J. Educational Sociology*, vol. 2, no. 4, pp. 188–197, 2019, doi: 10.29252/ijes.2.4.188.

[13] S. S. Nahrkhalaji, M. M. Shafiee, and S. Shafiee, “Organization design in motion: Designing an organization for agility,” *Proc. Design Society*, vol. 1, no. July, pp. 2349 - 2358, 2021, doi: 10.1017/pds.2021.496.

[14] N. Paethrangsi, V. Parnkan, and S. Teekasap, “Impacts of organization agility on the airline performance outcomes,” *E3S Web of Conferences*, vol. 389, no. 10, p. 05011, 2023, doi: 10.1051/e3sconf/202338905011.

[15] L. Aldianto, G. Anggadwita, A. Permatasari, I. R. Mirzanti, and I. O. Williamson, “Toward a business resilience framework for startups,” *Sustainability*, vol. 13, no. 6, p. 3132, 2021, doi: 10.3390/su13063132.

[16] M. M. Abrishamkar, Y. A. Abubakar, and J. Mitra, “The influence of workforce agility on high-growth firms: The mediating role of innovation,” *Int. J. Entrepreneurship and Innovation*, vol. 22, no. 3, pp. 146-160, 2020, doi: 10.1177/1465750320973896.

[17] C. Xue and J. Wang, “Proactive boundary-spanning search, organizational resilience, and radical green innovation,” *Business Strategy and the Environment*, vol. 32, no. 3, pp. 1834-1852, 2023, doi: 10.1002/bse.3576.

[18] A. Bouguerra, İ. Gölgeci, D. Gligor, Z. Khan, A. Arslan, and E. Tatoğlu, “Strategic agility and product development in emerging markets: The role of employee resilience and self-efficacy as microfoundations,” *IEEE Trans. Eng. Manage.*, vol. 71, no. November, pp. 1598 - 1612, 2024, doi: 10.1109/tem.2023.3333218.

[19] F. L. Oliva, M. H. Gomes Couto, R. F. Santos, and S. Bresciani, “The integration between knowledge management and dynamic capabilities in agile organizations,” *Management Decision*, vol. 57, no. 8, pp. 1960–1979, 2019, doi: 10.1108/md-06-2018-0670.

[20] E. J. Schwarz, P. Gregori, I. Krajger, and M. A. Wdowiak, “Entrepreneurial lean thinking for sustainable business modeling: A workshop design for incumbent firms,” *Nachhaltigkeitsmanagementforum | Sustainability Management Forum*, vol. 29, no. January, pp. 41-45, 2021, doi: 10.1007/s00550-020-00508-y.

[21] B. Baldassarre, G. Calabretta, N. Bocken, and T. Jaśkiewicz, “Bridging sustainable business model innovation and user-driven innovation: A process for sustainable value proposition design,” *J. Cleaner Production*, vol. 147, no. March, pp. 175–186, 2017, doi: 10.1016/j.jclepro.2017.01.081.

[22] V. Roblek, V. Dimovski, M. Meško, and J. Peterlin, “Evolution of organisational agility: A bibliometric study,” *Kybernetes*, vol. 51, no. 13, pp. 119–137, 2022, doi: 10.1108/k-11-2021-1137.

[23] A. Skala, “Sustainable transport and mobility–oriented innovative startups and business models,” *Sustainability*, vol. 14, no. 9, p. 5519, 2022, doi: 10.3390/su14095519.

[24] A. Ajanaku, “Artificial intelligence for startup risk and investment readiness assessment: A machine learning model from the African innovation ecosystem,” *Current Trends in Business Management*, vol. 3, no. 2, pp. 1–21, 2025, doi: 10.33140/ctbm.03.02.03.

[25] T. J. Oladuji, A. O. Akintobi, C. R. Nwangele, and O. Onifade, “A conceptual framework for strategic risk management in African venture capital: Combining market volatility with AI-driven data analysis,” *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 11, no. 3, pp. 635–673, 2024, doi: 10.32628/ijsrset241483.

[26] S. R. Chanthati, “A segmented approach to encouragement of entrepreneurship using data science,” *World J. Adv. Eng. Technol. Sci.*, vol. 12, no. 2, pp. 584–605, 2024, doi: 10.30574/wjaets.2024.12.2.0330.

[27] S. Ferro, D. Bottigliengo, D. Gregori, A. S. C. Fabricio, M. Gion, and I. Baldi, “Phenomapping of patients with primary breast cancer using machine learning-based unsupervised cluster analysis,” *J. Personalized Medicine*, vol. 11, no. 4, p. 272, 2021, doi: 10.3390/jpm11040272.