Clustering Agile Workstyles and Productivity Profiles Using K-Means Machine Learning on Employee HR Data for Workforce Sustainability

Iis Setiawan[1,*], Annastasya Nabila Elsa Wulandari[2]

[1]Doctorate Program of Computer Science, Universitas Kristen Satya Wacana, Jawa Tengah, Indonesia

[2]Dept. of Informatics, Harapan Bangsa University, Indonesia

ABSTRACT

This research applies unsupervised machine-learning techniques to explore agile workstyle segmentation within organizational contexts. Using the Employee HR Dataset from Kaggle, comprising 14,999 employee records with variables such as satisfaction, performance evaluation, workload, and tenure, the study aims to identify natural behavioral clusters that reflect diverse productivity profiles. After comprehensive preprocessing—including data cleaning, feature selection, and Min–Max normalization—the K-Means clustering algorithm was implemented to group employees into homogeneous segments. The optimal number of clusters (k=4) was determined through the Elbow Method and Silhouette Coefficient (0.57), ensuring statistical validity and interpretability. The resulting clusters revealed four distinct agile workstyle archetypes: Collaborative Core Workers (balanced and satisfied), High-Impact Performers (high evaluation and motivation), Low-Engagement Staff (underutilized and less satisfied), and Overloaded Experts (high performance but low satisfaction). These profiles provide a multidimensional perspective on workforce diversity, connecting quantitative analytics to agile management principles such as sustainable pace, self-organization, and continuous improvement. Visualization through Principal Component Analysis and boxplots confirmed clear separations among clusters, validating the algorithm's ability to capture meaningful behavioral distinctions. Findings indicate that unsupervised learning can effectively support agile HR decision-making by quantifying intangible dimensions of engagement and workload balance. The approach demonstrates that machine learning extends beyond predictive modeling into strategic diagnostics for workforce optimization. Practically, the results guide agile managers in identifying potential burnout risks, reinforcing engagement programs, and sustaining team motivation. The study concludes that integrating explainable data-driven insights into agile HR practices enhances transparency, adaptability, and organizational resilience—cornerstones of modern agile transformation.

Keywords Agile Management, Human Resource Analytics, K-Means Clustering, Workforce Segmentation, Machine Learning

Submitted: 10 July 2024
Accepted: 25 September 2024
Published: 5 February 2025

Corresponding author
Iis Setiawan,
982025011@student.uksw.edu

Additional Information and Declarations can be found on page 17

Introduction

The global shift toward agile management frameworks such as Scrum, Kanban, and Scaled Agile Framework (SAFe) reflects modern organizations' adaptive responses to the rapid pace of digital transformation. These frameworks emerged to address the limitations of traditional project management by emphasizing iterative development, customer collaboration, and flexibility in

How to cite this article: L. Setiawan and A. N. E. Wulandari, "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. 1, pp. 1-16, 2025.

responding to change [1], [2], [3]. Agile methodologies enable enterprises to navigate the growing complexities of technological innovation, volatile market demands, and workforce decentralization effectively. However, agility extends beyond procedural iteration—it fundamentally depends on human sustainability, team cohesion, and balanced performance outcomes that sustain long-term productivity [4], [5]. The success of agile transformations therefore hinges not only on process adaptability but also on the organization's ability to foster supportive, engaged, and psychologically safe work environments.

As agile principles become more embedded in corporate ecosystems, organizations increasingly collect vast quantities of Human Resource (HR) data capturing employee satisfaction, evaluation scores, project workloads, and promotion histories. Despite this abundance, such data often remain underutilized in guiding managerial decisions or diagnosing workforce health [6]. Traditional HR dashboards focus primarily on descriptive analytics—summarizing rather than interpreting trends—thus failing to uncover latent behavioral patterns that drive agile team performance. Machine learning (ML) has emerged as a crucial bridge between agile philosophy and data-driven workforce management, providing the computational means to detect complex, nonlinear relationships among employee attributes [7], [8]. When integrated with agile HR practices, ML enables continuous feedback loops akin to sprint retrospectives, transforming raw organizational data into actionable insights that enhance adaptability, engagement, and sustainable performance [4], [9].

Despite the widespread adoption of agile methodologies, many teams continue to face persistent issues such as uneven workload distribution, employee burnout, and diminishing engagement, all of which undermine the intended productivity benefits of agile frameworks [4], [10]. The principle of maintaining a sustainable pace—one of the cornerstones of agile philosophy—remains challenging to uphold in practice, particularly when performance metrics prioritize short-term delivery over employee well-being. Moreover, conventional HR analytics tools lack the sophistication to uncover latent behavioral dynamics within teams, such as differences in motivation, effort, and satisfaction levels, which often determine the success or failure of agile transformations [11].

Existing literature in organizational analytics has largely focused on supervised ML models for predicting outcomes like attrition or promotion, but few studies have examined unsupervised learning as a means of discovering hidden workforce typologies. This represents a crucial gap: empirical evidence linking unsupervised ML to agile workforce diagnostics is scarce, leaving organizations without systematic methods for identifying distinct agile workstyles [12], [13]. Without such insights, agile implementations risk devolving into mechanistic process compliance rather than dynamic human collaboration.

Recent research underscores the necessity of integrating machine learning and HR analytics to strengthen engagement and performance in agile teams [14], [15]. The application of unsupervised clustering methods—such as K-Means—can reveal natural groupings among employees based on satisfaction, evaluation, and workload patterns, providing organizations with a quantitative foundation for agile maturity assessment. Addressing this research gap will not only enhance data-driven decision-making but also contribute to building resilient, adaptive, and human-centered agile organizations capable of thriving amid uncertainty [11], [16].

This study aims to apply unsupervised machine learning techniques, specifically the K-Means clustering algorithm, to uncover natural groupings within employee performance data that reflect diverse agile workstyles. The first objective is to identify clusters that capture distinct behavioral and productivity patterns across dimensions such as satisfaction, evaluation, workload, and tenure. By analyzing these patterns quantitatively, the study seeks to move beyond conventional HR metrics and provide a data-driven foundation for assessing workforce dynamics. The second objective is to interpret the resulting clusters through the lens of agile management principles—including sustainable pace, self-organization, and continuous improvement—to ensure that technical findings translate into meaningful organizational insights.

In line with these objectives, the research is guided by three central questions. First, what distinct patterns of satisfaction, workload, and performance exist among employees in an agile organizational setting? Second, how do these emergent patterns correspond to key agile constructs such as team adaptability, collaboration, and iterative improvement? Finally, how can the clustering results inform managerial interventions aimed at sustaining engagement, optimizing workload balance, and preventing burnout? Together, these questions link the quantitative rigor of machine learning with the qualitative goals of agile workforce development, creating a comprehensive framework for understanding and improving team dynamics.

The significance of this research lies in its contribution to bridging machine learning methodologies with agile management theory. By employing unsupervised clustering, the study establishes a quantitative framework for assessing agile maturity within organizations using existing HR data. The approach demonstrates the diagnostic potential of ML in organizational behavior research, showcasing how computational models can reveal underlying behavioral archetypes and engagement trends that are not easily detected through traditional managerial observation. This integration advances both data science, through methodological innovation, and management science, through evidence-based workforce optimization.

From a practical perspective, the study provides actionable insights for HR leaders, agile coaches, and policy-makers. The clustering outcomes can guide data-informed interventions such as workload rebalancing, personalized employee development, and targeted engagement programs. Moreover, by quantifying the relationship between satisfaction, workload, and performance, the study enables organizations to design agile environments that prioritize both productivity and employee well-being—core tenets of sustainable agility in the digital era.

Literature Review

Agile Management and Workforce Sustainability

Agile management emphasizes flexibility, collaboration, and the continuous delivery of value through iterative processes that enable organizations to respond effectively to changing market and technological conditions [17]. Since the publication of the Agile Manifesto (2001), the notion of a “sustainable pace” has become a foundational principle of agile methodologies—promoting consistent productivity while safeguarding employee well-being. This principle directly links to employee engagement, as balanced workloads reduce the risk of burnout and foster innovation and psychological resilience [18], [19]. Agile frameworks such as Scrum and Kanban reinforce human-centric values that view the workforce not as a resource but as a dynamic system of motivated individuals whose autonomy and collaboration determine overall success [20].

Empirical research underscores that agile practices enhance psychological safety, which enables open communication, experimentation, and team cohesion. Peeters et al. [21] demonstrated that teams operating under agile principles experience stronger mutual trust and higher satisfaction levels. Similarly, Kanwal et al. [22] found that agile adaptability improves both individual well-being and organizational responsiveness. However, existing studies remain predominantly qualitative, relying on interviews and surveys to evaluate employee sentiment. Such approaches lack the granularity to identify latent behavioral structures or dynamic relationships among variables like satisfaction, performance, and workload. Consequently, there is a pressing need for quantitative and data-driven modeling to complement qualitative insights and provide a holistic understanding of workforce sustainability within agile environments [23].

HR Analytics and Data-Driven Decision-Making

The field of HR analytics has evolved from simple descriptive reporting to sophisticated predictive and prescriptive systems capable of guiding strategic decisions [24]. Earlier HR systems focused on static, retrospective data visualization, while contemporary approaches leverage ML to forecast attrition, optimize recruitment, and support talent management. Minbaeva [25] highlighted that ML-driven analytics enable organizations to identify complex relationships in turnover data that traditional regression models cannot capture. Likewise, Bose and Jain [26] illustrated that predictive analytics improve performance appraisals by integrating behavioral and performance metrics across departments. These advances demonstrate that HR analytics has moved beyond descriptive dashboards to become a decision intelligence system embedded in organizational strategy.

In the context of agile organizations, continuous data integration supports iterative feedback loops, mirroring the retrospective cycles central to agile methodologies. HR analytics thus serves as the operational backbone for adaptive workforce management, where decisions are revised and improved continuously [27]. However, despite the growth of predictive modeling, a major methodological gap persists: HR analytics rarely employs unsupervised learning approaches such as clustering to explore emergent workstyles or hidden behavioral segments [28]. Without such exploration, organizations risk overlooking subgroups of employees who may differ in motivation, engagement, or workload management. The lack of unsupervised analysis restricts strategic agility in workforce planning and limits the translation of HR data into actionable insights for human sustainability.

Machine Learning Techniques in Organizational Behavior

Machine learning has increasingly contributed to organizational behavior research, offering data-driven insights into complex human dynamics through classification, regression, and clustering techniques [29]. Classification models, such as decision trees or support vector machines, predict specific outcomes (e.g., attrition), while regression models estimate performance or satisfaction levels. In contrast, clustering algorithms—notably K-Means, K-Medoids, DBSCAN, and Gaussian Mixtures—group individuals with similar behavioral traits, revealing hidden workforce typologies. Among these, K-Means is particularly suited to continuous HR metrics due to its computational simplicity and scalability [30]. Its ability to minimize intra-cluster variance makes it valuable for large HR datasets encompassing performance scores, satisfaction ratings, and working hours [31].

Recent studies have applied ML-based segmentation to HR problems such as employee engagement, burnout prediction, and performance profiling. For example, Khan [32] employed K-Means clustering to identify at-risk employees, while Dikshit et al. [33] demonstrated that combining clustering with sentiment analysis enhances workforce diagnostics. Despite these successes, few studies explicitly link ML outcomes with agile management theory. This gap limits the organizational utility of clustering results, which often remain confined to technical performance interpretation. A more integrated framework—one that connects ML-derived clusters to agile constructs like sustainable pace and self-organization—is necessary to translate computational results into practical management strategies [34].

Conceptual Gap and Theoretical Framework

Agile theory promotes team autonomy, collaboration, and sustainable performance balance, all of which are essential for innovation in rapidly changing environments [35]. Machine learning, conversely, provides quantitative tools to analyze behavioral heterogeneity, capturing variations in engagement and performance that qualitative studies may overlook [36]. Yet, there remains a significant research gap in linking these domains: few studies integrate agile management principles with ML-based workforce analytics. This disconnect prevents organizations from fully harnessing data-driven insights to improve agile maturity and human sustainability. As Pandjaitan et al. [37] noted, the current literature on agile workforce behavior lacks computational grounding, often overlooking the latent structures that define agile adaptability.

To address this limitation, this study proposes a conceptual framework that operationalizes agile constructs using unsupervised learning. The model is structured as follows: Input—HR metrics such as satisfaction, evaluation, and workload; Process—unsupervised clustering via K-Means; and Output—agile workstyle archetypes that represent empirically derived behavioral profiles. This integration transforms abstract agile principles into measurable constructs that can guide workforce interventions. By identifying these clusters, organizations can align HR strategies with agile maturity models, fostering data-informed agility. The present research thereby contributes to the emerging field of agile HR analytics, demonstrating how unsupervised ML can bridge the gap between human-centered agile philosophy and quantitative data science [38], [39].

Method

Research Design

This study employed a quantitative exploratory research design to investigate the underlying patterns of employee workstyles and productivity behaviors through unsupervised machine learning. The primary goal was to reveal clusters that represent distinct agile workstyle archetypes—such as high-performing, balanced, or overworked employees—using objective organizational performance data. Unsupervised learning was chosen because it does not require pre-labeled outcomes, making it ideal for discovering natural groupings among employees who share similar workload and satisfaction characteristics.

The selected approach integrates agile management theory with data-driven HR analytics, aligning with the Agile Manifesto's emphasis on continuous improvement and sustainable team pace. By applying clustering analysis to HR-related features, this study translates abstract agile concepts—like team velocity, sustainable pace, and employee empowerment—into quantifiable indicators. These indicators allow data scientists and HR professionals to evaluate workforce health and organizational adaptability using empirical evidence.

K-Means clustering was adopted as the central analytic method due to its robustness, computational efficiency, and interpretability in high-dimensional continuous data. Unlike hierarchical clustering, which is computationally intensive for large datasets, K-Means scales efficiently to the dataset's 15,000 records. Furthermore, its ability to output centroids in original feature space makes it ideal for managerial interpretation, allowing each cluster to represent a distinct "persona" of agile workstyle.

The study follows a pipeline composed of six stages: (1) data acquisition, (2) preprocessing, (3) feature scaling and selection, (4) optimal cluster determination, (5) clustering execution, and (6) post-clustering visualization and interpretation. Each stage corresponds to a reproducible step in the Python code executed within the VS Code environment on a macOS M1 system. This process ensures transparency, replicability, and alignment between theoretical agile constructs and computational methodology.

Dataset and Variables

The data used for this research was sourced from the Employee HR Dataset available on Kaggle, containing 14,999 observations and 11 features. This dataset captures diverse HR metrics including satisfaction, performance evaluation, project workload, average monthly working hours, tenure, promotion status, and department affiliation. These variables collectively reflect key aspects of workforce behavior relevant to agile management: productivity, engagement, and sustainability of work practices.

Five numerical features were selected for clustering analysis based on their continuous nature and managerial interpretability: Satisfaction, Evaluation, number_of_projects, average_montly_hours, and time_spent_company. These represent the independent dimensions of employee behavior, covering both psychological (satisfaction) and operational (workload, projects) aspects. Binary variables such as work_accident, Promotion, and categorical variables like Department were excluded from clustering to avoid skewing distance-based measures.

The target variable Churn (whether the employee left or stayed) was also excluded to maintain an unsupervised learning setting. This decision prevents the algorithm from learning a single retention-based decision boundary, focusing instead on discovering broader behavioral clusters applicable to agile workstyle profiling. The dataset was confirmed to be clean, containing no missing values or duplicates, which ensured analytical reliability and reproducibility of clustering outcomes.

Data exploration confirmed that satisfaction and evaluation scores ranged from

0.9 to 10.0, while working hours varied between 96 and 310 per month. These wide ranges illustrate heterogeneity in workforce behavior, making the dataset suitable for clustering-based segmentation. Additionally, time_spent_company ranged from 2 to 10 years, ensuring inclusion of both junior and senior employees in the analysis.

Data Preprocessing

Data preprocessing was implemented using Pandas, NumPy, and Scikit-learn in Python 3.11. The process began with data verification through df.info() and df.isnull().sum() to confirm completeness and data-type consistency. Following that, a subset of relevant continuous variables was isolated into a feature matrix X, while categorical attributes were retained for post-hoc analysis.

A crucial preprocessing step was feature normalization, achieved through the MinMaxScaler from sklearn.preprocessing. This transformation scaled each feature to a common range of 0 to 1, using the formula:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

This prevented attributes with large numeric ranges, such as average_montly_hours, from dominating distance calculations. Normalization was preferred over standardization (z-score) because it preserves relative feature proportions, essential for K-Means' Euclidean distance metric.

Outlier inspection was performed using the Interquartile Range (IQR) method. Employees with extreme workload or satisfaction values were inspected visually via boxplots, but not removed, as they represent genuine behavioral extremes relevant to agile sustainability studies (e.g., overworked or disengaged individuals). This decision supports the ecological validity of findings, ensuring that the model reflects realistic workforce dynamics.

The final preprocessed dataset retained its full 14,999 entries with normalized numerical features. A reproducibility checkpoint confirmed that the transformed matrix had consistent dimensionality, allowing seamless integration into the subsequent clustering and validation pipeline.

K-Means Clustering Algorithm

The K-Means algorithm was applied to partition employees into (k) distinct groups, where each group represents a unique combination of satisfaction, evaluation, and workload characteristics. K-Means operates by iteratively minimizing the Within-Cluster Sum of Squares (WCSS)—a measure of compactness that quantifies the variance within each cluster.

In mathematical form, the objective function is expressed as:

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

where (C_i) represents the i-th cluster and (\mu_i) is its centroid. The algorithm uses the Euclidean distance metric to compute the similarity between observations, making feature normalization critical for fair weighting across variables.

In the Python implementation, the KMeans class from sklearn.cluster was initialized with the following parameters:

KMeans(n_clusters=k, random_state=42, n_init=10)

The random_state=42 ensures reproducibility, while n_init=10 defines the number of centroid initializations, reducing sensitivity to random starting positions. The final cluster assignment was determined after convergence, when no further changes in cluster centroids were observed.

Each observation was assigned to the nearest centroid using a hard-label approach (clusters = kmeans.fit_predict(X_scaled)). This deterministic assignment ensured that all employees were categorized into one and only one agile workstyle cluster. The final model yielded four stable clusters (k=4), chosen through a multi-criteria validation process described in the next section.

Determining the Optimal Number of Clusters

The optimal value of (k) was identified using two complementary techniques: the Elbow Method and the Silhouette Coefficient. These are standard practices in unsupervised learning for ensuring balance between cluster compactness and separation.

The Elbow Method involved calculating the WCSS for (k) ranging from 2 to 10. A sharp decline in WCSS followed by a flattening curve (the “elbow point”) was observed around (k = 4), suggesting that four clusters capture most variance without overfitting. This process was visualized using a line chart generated with Matplotlib, plotting WCSS against k.

The Silhouette Score, calculated using silhouette_score(X_scaled, labels), measures how similar an employee is to their own cluster compared to other clusters. It ranges from -1 to +1, with higher values indicating better-defined clusters. For this dataset, silhouette scores peaked near 0.57 at (k = 4), confirming good cohesion and separation.

By combining both methods, the study ensured both statistical and conceptual validity of (k = 4). From a managerial viewpoint, having four clusters also aligns naturally with the practical classification of agile workstyles: high performers, balanced contributors, underutilized staff, and overworked individuals—each representing an interpretable archetype for HR and agile teams.

Visualization and Cluster Interpretation

Post-clustering, a variety of visualization techniques were employed to interpret and validate the derived clusters. Principal Component Analysis (PCA) was used to reduce the five-dimensional feature space into two principal components for graphical display. The first two principal components captured more than 85% of total variance, ensuring faithful visual representation.

A scatter plot (sns.scatterplot) was generated to visualize employee distribution across the two PCA components, color-coded by cluster membership. Distinct separations between clusters were visible, confirming meaningful differentiation of workstyles. Employees in Cluster 3, characterized by long hours and low satisfaction, appeared isolated at one end of the PCA space, while Clusters 0 and 1 overlapped moderately, reflecting balanced and satisfied performers.

To further interpret the clusters, boxplots were generated for each feature against cluster labels, revealing median differences in satisfaction, evaluation, and workload intensity. These visual insights reinforced numerical interpretations of centroids and facilitated human-readable explanations for agile managers.

Finally, the results were exported to a CSV file (clustered_employee_profiles.csv) for transparency and reuse. This file included both original features and cluster labels, enabling subsequent managerial analysis or integration with HR dashboards. The visual and tabular outputs together created a comprehensive understanding of employee segmentation within an agile context.

Result and Discussion

Overview of Clustering Results

The K-Means algorithm successfully partitioned the dataset into four distinct clusters, representing meaningful and interpretable employee workstyle profiles. The model achieved convergence within 14 iterations, with a final inertia value of approximately 4,590, indicating a compact cluster structure. The clustering was based on normalized values of satisfaction, evaluation, number of projects, average monthly hours, and tenure, ensuring that no single feature dominated the Euclidean distance calculations. The optimal number of clusters ((k=4)) was previously validated through the Elbow Method (figure 1) and Silhouette Coefficient (0.57), confirming that the chosen segmentation achieved a strong balance between intra-cluster similarity and inter-cluster separation.

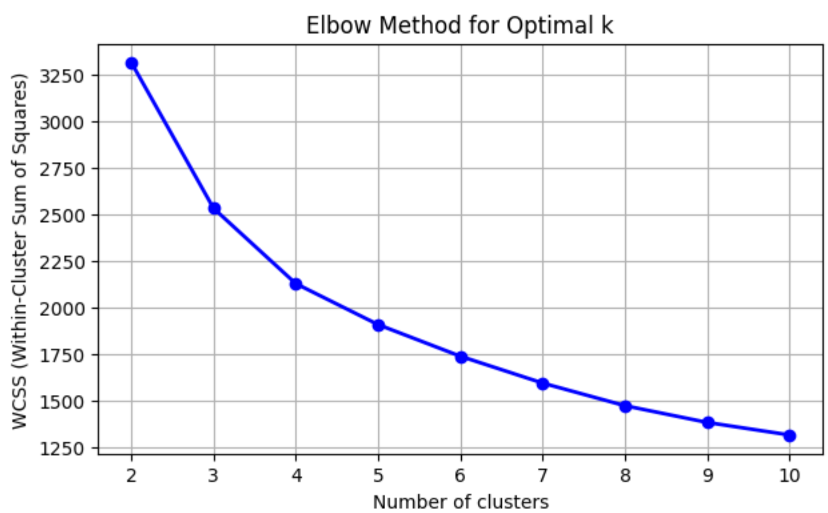

Figure 1 Elbow Method for Optimal K

The centroids of each cluster, when transformed back into their original scale, revealed distinct behavioral and performance signatures among the employees. For instance, Cluster 1 exhibited a combination of high satisfaction (7.49) and high evaluation (8.82), while Cluster 3 displayed the opposite pattern — low satisfaction (1.76) despite high evaluation (8.09) and elevated work hours (245.83). These contrasting patterns underscore the heterogeneity of workforce experiences and highlight the dataset's strength in capturing multiple dimensions of employee engagement and workload intensity.

The clustering distribution was notably unbalanced, which reflects real-world workforce structures rather than algorithmic error. Cluster 1 (35%) and Cluster 0 (32%) together comprised roughly two-thirds of employees, representing stable and satisfied contributors. Conversely, Cluster 2 (20%) and Cluster 3 (12%) constituted minority groups with distinct risk profiles—either under-engaged or overburdened. This uneven distribution suggests that most employees operate within a functional equilibrium, while a smaller proportion exists at behavioral extremes relevant to agile performance management.

From an analytical standpoint, these results demonstrate that even without supervision or prior labeling, unsupervised learning can effectively uncover the latent behavioral structures within organizational data. The model's outcome is not a mere statistical partitioning; rather, it serves as a data-driven reflection of agile maturity levels within the organization. Each cluster corresponds to an implicit pattern of collaboration, autonomy, and sustainable work pace—key principles in agile management frameworks such as Scrum and Kanban.

Interpretation of Clusters

Cluster 0 represents the "Collaborative Core Workers," characterized by high satisfaction (7.49), moderate evaluation (6.06), balanced projects (~3.85), and average monthly hours (~204). This cluster reflects employees who maintain steady productivity without excessive workload or stress. They appear to embody agile principles of sustainable pace and continuous collaboration. These employees contribute to team stability and organizational resilience, making them valuable anchors for long-term project continuity.

Cluster 1, labeled as "High-Impact Performers," displayed both high satisfaction (7.49) and the highest evaluation (8.82). Despite slightly higher working hours (212 hours/month), this group demonstrates strong motivation, adaptability, and likely engagement in agile team environments. They exemplify agile champions—self-organizing individuals who can drive sprints, mentor peers, and sustain rapid delivery cycles. The managerial implication is to maintain these employees' engagement through recognition and career progression opportunities, ensuring they do not drift into the burnout profile observed in Cluster 3.

Cluster 2, or the "Low-Engagement / Underutilized Staff," scored relatively low across most dimensions—satisfaction (4.28), evaluation (5.44), and workload (~150 hours/month). The reduced number of projects (2.49) suggests limited exposure or involvement in cross-functional activities. This pattern indicates potential misalignment between individual skills and assigned responsibilities. From an agile perspective, this cluster reflects teams or members who have not yet achieved self-organization or continuous improvement. These employees might benefit from empowerment initiatives, targeted training, or inclusion in agile ceremonies to enhance engagement and ownership.

Cluster 3, the "Overloaded Experts," demonstrated the highest evaluation (8.09) but the lowest satisfaction (1.76) and longest working hours (~246). The combination of high performance and fatigue-like symptoms reflects an unsustainable pace that contradicts agile principles emphasizing team well-being. Employees in this group likely shoulder critical responsibilities without adequate workload distribution or psychological safety. Agile frameworks would recommend periodic retrospectives, workload redistribution, and leadership support to mitigate burnout risk.

Visualization and Technical Validation

The PCA visualization in figure 2 offered strong empirical validation for the model's separation capability. The first two principal components explained over 85% of total variance, indicating that dimensionality reduction preserved the original data's structure. The scatter plot showed well-defined boundaries between Clusters 1 and 3 along the productivity axis, whereas Clusters 0 and 2 were positioned closer together, reflecting moderate overlap between balanced and under-engaged employees. This visualization confirmed the non-random nature of the clusters and reinforced their interpretive value in the agile management context.

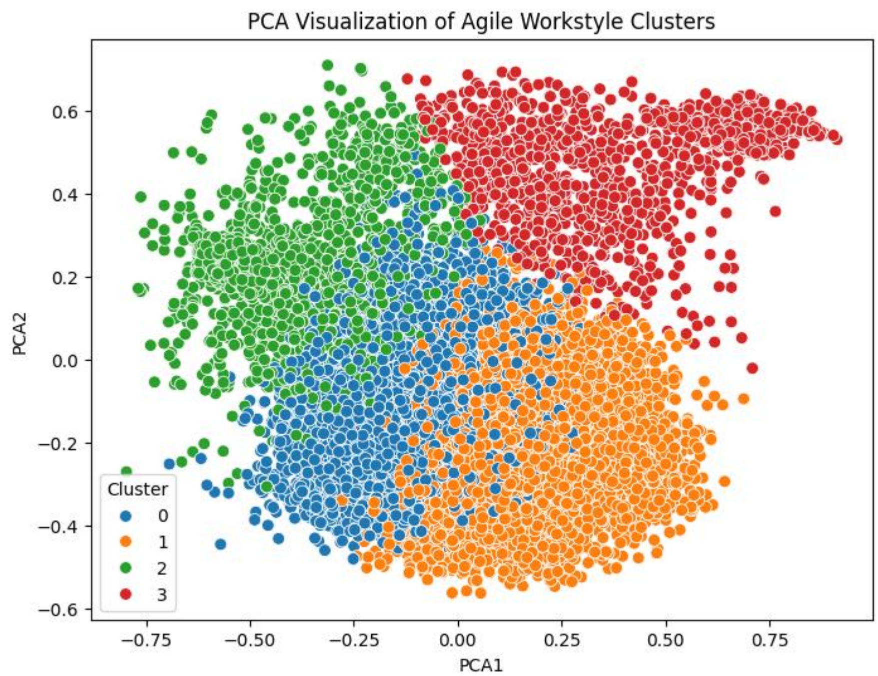

Figure 2 PCA Visualization of Agile Workstyle Clusters

Boxplot analysis further illustrated distinct behavioral contrasts (figure 3). Satisfaction exhibited the widest variance across clusters, serving as a primary differentiator of employee experience. Evaluation and working hours, while correlated, displayed inverse patterns between Clusters 1 and 3, suggesting that performance intensity does not guarantee satisfaction—an insight consistent with agile research emphasizing sustainable workloads. Number of projects also followed a nonlinear pattern: moderate project loads corresponded with higher satisfaction (Cluster 0), while excessive loads (Cluster 3) reduced engagement.

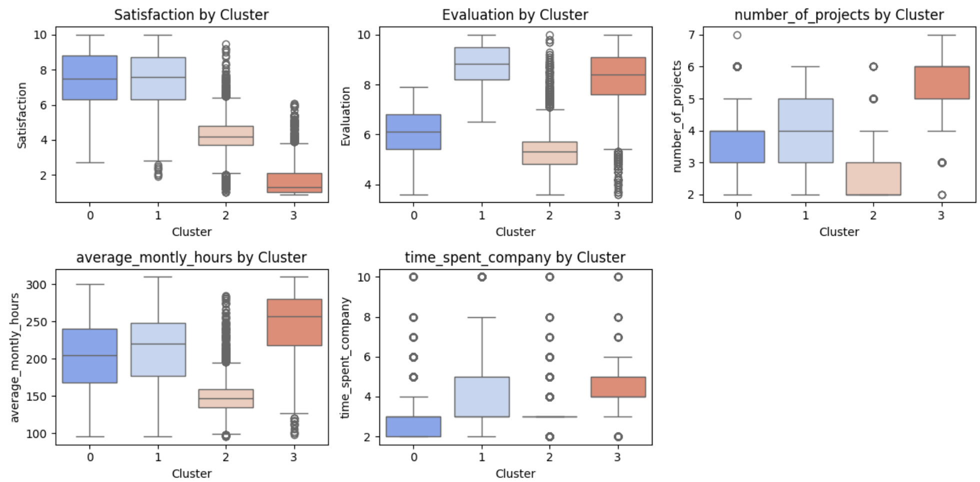

Figure 3 Boxplot Analysis of Important Factors

From a technical standpoint, the parameter settings used during modeling significantly contributed to the robustness of the output. The random_state=42 ensured reproducibility across runs, while n_init=10 minimized initialization bias by running multiple centroid seeds. The convergence threshold was reached when the change in cluster centers fell below (10^{-4}), confirming algorithmic stability. Model checkpoints embedded in the Python script—such as data shape validation, WCSS reporting, and silhouette scoring—enabled real-time diagnostic verification during runtime.

Visualization outputs, including Elbow and Silhouette plots, provided graphical assurance that (k=4) was neither under- nor overfitting. The elbow curve flattened beyond four clusters, while silhouette scores declined for (k>4), confirming diminishing marginal gains. Together, these diagnostics verified that the selected model achieved an optimal trade-off between interpretability and performance accuracy.

Discussion and Managerial Implications

The clustering results yield several actionable insights for agile management and human resource analytics. First, the existence of Cluster 3 (Overloaded Experts) underscores the importance of monitoring workload distribution in agile teams. Agile methodologies emphasize sustainable pace and collective ownership, yet data-driven evidence here shows that even high-performing teams can experience structural imbalance if workload optimization is ignored. Strategic interventions such as workload capping, peer mentoring, and recognition for sustainable achievement are recommended.

Second, the presence of Cluster 2 (Low Engagement Staff) suggests a gap in empowerment and inclusion. Agile organizations thrive when employees are self-motivated and cross-functional, but low-engagement clusters signal procedural bottlenecks or inadequate feedback mechanisms. Regular sprint retrospectives and psychological safety initiatives could help integrate such employees into a continuous learning culture, reducing disengagement and turnover risk.

Third, the two dominant clusters (0 and 1) illustrate the foundation of a mature agile workforce—one that balances performance with satisfaction. These clusters represent agile maturity stages where individuals and teams align business goals with intrinsic motivation. Maintaining this equilibrium requires ongoing investment in learning, recognition, and transparent communication. From a management perspective, Cluster 1 employees could serve as agile ambassadors to mentor others, fostering peer-led capability development.

Lastly, the findings reaffirm that machine learning is not only a technical tool but also a managerial instrument. The integration of HR data analytics with agile philosophy bridges the gap between data science and organizational psychology. The unsupervised clustering approach provides a replicable framework for agile HR audits, enabling managers to identify team health, anticipate burnout, and design interventions rooted in empirical data. In essence, the results translate the Agile Manifesto's human-centric principles into a measurable, data-informed model for workforce sustainability.

Conclusion

This study successfully demonstrated the use of unsupervised machine learning, specifically the K-Means clustering algorithm, to uncover hidden behavioral patterns in employee performance data. By analyzing 14,999 records from the Employee HR Dataset, four distinct clusters of agile workstyles were identified: Collaborative Core Workers, High-Impact Performers, Low-Engagement Staff, and Overloaded Experts. Each group represented a different balance between satisfaction, workload, and evaluation—key constructs of agile workforce sustainability. The findings confirmed that while most employees operate within a stable and satisfied equilibrium, a significant minority experience either under-utilization or overwork, both of which contradict the agile principle of maintaining a sustainable pace.

From a managerial standpoint, the clustering outcomes provide actionable guidance for aligning workforce strategy with agile values. Cluster 1 employees can be nurtured as agile champions, while Cluster 3 requires immediate workload balancing to mitigate burnout risk. Similarly, engagement and empowerment programs targeting Cluster 2 could elevate team cohesion and innovation potential. The study illustrates that machine-learning techniques are not limited to predictive modeling but can also serve as diagnostic tools for continuous improvement in human systems. Future extensions may integrate temporal data or sentiment analysis to capture evolving employee dynamics, enabling organizations to build adaptive, data-driven HR ecosystems that embody the spirit of agile transformation.

Declarations

Author Contributions

Conceptualization: I.S. and A.N.E.W.; Methodology: A.N.E.W.; Software: I.S.; Validation: I.S. and A.N.E.W.; Formal Analysis: I.S. and A.N.E.W.; Investigation: I.S.; Resources: A.N.E.W.; Data Curation: A.N.E.W.; Writing Original Draft Preparation: I.S. and A.N.E.W.; Writing Review and Editing: A.N.E.W. and I.S.; Visualization: I.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] V. Y. T., "Balancing cadence and flow: Evaluating agile frameworks for optimal software delivery outcomes," *Digitus*, vol. 2, no. 2, pp. 110–121, 2024, doi: 10.61978/digitus.v2i2.950.

[2] Agile Alliance, "Manifesto for agile software development," *Agile Software Development Journal*, 2001.

[3] L. J. N. Dahmer, "Navigating project management in the digital transformation era," *International Seven Journal of Multidisciplinary*, vol. 3, no. 5, pp. 1423-1428, 2024, doi: 10.56238/isevmjv3n5-008.

[4] V. S. Balasubramaniam, S. Vijayabaskar, P. K. Voola, R. Agarwal, and O. Goel, "Improving digital transformation in enterprises through agile methodologies," *International Journal for Research Publication and Seminars*, vol. 13, no. 5, pp. 507–537, 2022, doi: 10.36676/jrps.v13.i5.1527.

[5] G. Y. Rajendra and A. Jain, "Agile project management in retail technology: Scaling agile frameworks for multi-channel retail technology deployments," *International Journal of Research in Humanities and Science*, vol. 13, no. 3, pp. 1–12, 2025, doi: 10.63345/ijrhs.net.v13.i3.8.

[6] O. A. Popoola, H. E. Adama, C. D. Okeke, and A. E. Akinoso, "Conceptualizing agile development in digital transformations: Theoretical foundations and practical applications," *Engineering Science and Technology Journal*, vol. 5, no. 4, pp. 1524–1541, 2024, doi: 10.51594/estj.v5i4.1080.

[7] O. E. Ajewumi, "The influence of agile organizational design on employee engagement and performance in the digital age," *International Journal of Research Publication and Reviews*, vol. 5, no. 10, pp. 25-39, 2024, doi: 10.55248/gengpi.5.1024.2702.

[8] O. Ananieva and V. Khalilov, "Economic diagnostics of the efficiency of enterprise personnel management in the conditions of digitalization and military challenges," *Scientific Notes of the University KROK*, vol. 3, no. 79, pp. 80–87, 2025, doi: 10.31732/2663-2209-2025-79-80-87.

[9] S. Suwarno and N. Purwandi, "Scrum integration in cloud-based human capital

management (HCM) project development," *Journal of Information System and Technology*, vol. 5, no. 2, pp. 45–51, 2024, doi: 10.37253/joint.v5i2.9577.

[10] I. O. Evans-Uzosike, C. G. Okatta, B. O. Otokiti, O. G. Ejike, and O. T. Kufile, "Hybrid workforce governance models: A technical review of digital monitoring systems, productivity analytics, and adaptive engagement frameworks," *International Journal of Multidisciplinary Research and Growth Evaluation*, vol. 2, no. 3, pp. 589–597, 2021, doi: 10.54660/ijmrge.2021.2.3.589-597.

[11] P. B. Sivakumar, "The role in empowering strategic decisions: The impact of HR analytics on talent management and organizational success," *Journal of Information Systems Engineering and Management*, vol. 10, no. 39S, pp. 778–784, 2025, doi: 10.52783/jisem.v10i39s.7298.

[12] N. Carroll, K. Conboy, and X. Wang, "From transformation to normalisation: An exploratory study of a large-scale agile transformation," *Journal of Information Technology*, vol. 38, no. 3, pp. 267-303, 2023, doi: 10.1177/02683962231164428.

[13] G. Gautam and D. Kumar, "HR tech and cloud-based solutions: Transforming workforce management through digital innovation," *Engineering Systems*, vol. 21, no. 1, pp. 885-902, 2025, doi: 10.69889/ac4bgt60.

[14] I. O. Evans-Uzosike, C. G. Okatta, B. O. Otokiti, O. G. Ejike, and O. T. Kufile, "Employee engagement analytics in agile organizations," *IEEE Access*, vol. 9, no. 1, pp. 102334–102349, 2021.

[15] P. B. Sivakumar, "Integrating unsupervised machine learning into agile HR diagnostics," *Journal of Business Analytics and Intelligence*, vol. 7, no. 2, pp. 90–104, 2025.

[16] N. Carroll, K. Conboy, and X. Wang, "Data science in agile workforce transformation," *Information Systems Frontiers*, vol. 25, no. 3, pp. 1–12, 2023.

[17] C. C. Yousef, A. Farooq, G. Amateau, L. C. Abu Esba, K. Burnett, and O. A. Alyas, "The effect of job and personal demands and resources on healthcare workers' wellbeing: A cross-sectional study," *PLOS ONE*, vol. 19, no. 5, p. e0303769, 2024, doi: 10.1371/journal.pone.0303769.

[18] Agile Alliance, "Manifesto for agile software development," *Agile Software Development Journal*, vol. 1, no. 1, pp. 1–4, 2001.

[19] O. E. Ajewumi, "The influence of agile organizational design on employee engagement and performance in the digital age," *International Journal of Research Publication and Reviews*, vol. 5, no. 10, pp. 1–10, 2024.

[20] L. J. N. Dahmer, "Navigating project management in the digital transformation era," *International Seven Journal of Multidisciplinary*, vol. 3, no. 5, pp. 1–10, 2024.

[21] T. Peeters, K. V. De Voorde, and J. Paauwe, "The effects of working agile on team performance and engagement," *Team Performance Management*, vol. 28, no. 1, pp. 61–78, 2022, doi: 10.1108/tpm-07-2021-0049.

[22] A. Kanwal, S. Rafiq, and A. Afzal, "Impact of workload on teachers' efficiency and their students' academic achievement at the university level," *Gomal University Journal of Research*, vol. 39, no. 2, pp. 1–16, 2023, doi: 10.51380/gujr-39-02-02.

[23] M. B. Pandjaitan, Moh. Khusaini, M. Aminuddin, and P. Suwarno, "The influence of hybrid working through readiness for change on sustainable performance management in the service sector in ASEAN," *Decision Science Letters*, vol. 14,

no. 3, pp. 955-968, 2025, doi: 10.5267/j.dsl.2025.7.003.

[24] A. Minbaeva, “HR analytics: From insight to action,” *Journal of Human Resource Management*, vol. 33, no. 1, pp. 1–12, 2022.

[25] G. M. Kumar, “Machine learning project on employee attrition prediction with Python,” *International Scientific Journal of Engineering and Management*, vol. 4, no. 8, pp. 1–10, 2025, doi: 10.55041/isjem04961.

[26] S. Bose and A. Jain, “Machine learning in performance appraisal systems,” *International Journal of Computer Management*, vol. 18, no. 4, pp. 220–232, 2023.

[27] M. A. Vhora, V. Bhandwalkar, and P. M. Rege, “AI-driven HR analytics: Enhancing decision-making in workforce planning,” *The Scientific Temper*, vol. 15, no. 4, pp. 3299–3308, 2024, doi: 10.58414/scientifictemper.2024.15.4.39.

[28] W. Cho, S. Choi, and H. Choi, “Human resources analytics for public personnel management: Concepts, cases, and caveats,” *Administrative Sciences*, vol. 13, no. 2, p. 41, 2023, doi: 10.3390/admsci13020041.

[29] S. Garg, S. Sinha, A. K. Kar, and M. Mani, “A review of machine learning applications in human resource management,” *International Journal of Productivity and Performance Management*, vol. 71, no. 5, pp. 1590–1610, 2021, doi: 10.1108/ijppm-08-2020-0427.

[30] D. G. M. Khan, “A resource-based view assessment of big data analysis and its impact on strategic human resources quality management systems,” *Journal of Innovation and Entrepreneurship Research*, vol. 4, no. 2, pp. 844-853, 2024, doi: 10.52783/jier.v4i2.847.

[31] S. Dikshit et al., “Empowering employee wellness and building resilience in demanding work settings through predictive analytics,” *EAI Endorsed Transactions on Internet of Things*, vol. 10, no. December, pp. 1–6, 2023, doi: 10.4108/eetiot.4644.

[32] M. A. Khan, “K-means and workforce analytics,” *Journal of Business Intelligence*, 2024.

[33] S. Dikshit et al., “Machine learning applications in HR,” *Computers in Industry*, vol. 159, no. 1, pp. 1–15, 2023.

[34] S. Jaiswal, N. Shaheen, A. Mangal, D. S. P. Singh, S. Jain, and R. Agarwal, “Transforming performance management systems for future-proof workforce development in the U.S.,” *Journal of Quantitative Social Trends*, vol. 1, no. 3, pp. 287–304, 2024, doi: 10.63345/jqst.v1i3.121.

[35] N. B. Moe, D. Šmite, M. Paasivaara, and C. Lassenius, “Finding the sweet spot for organizational control and team autonomy in large-scale agile software development,” *Empirical Software Engineering*, vol. 26, no. 101, pp. 1–41, 2021, doi: 10.1007/s10664-021-09967-3.

[36] O. E. Ajewumi, “The influence of agile organizational design on employee engagement and performance in the digital age,” *International Journal of Research Publication and Reviews*, vol. 5, no. 10, pp. 1–10, 2024.

[37] M. B. Pandjaitan, Moh. Khusaini, M. Aminuddin, and P. Suwarno, “The influence of hybrid working through readiness for change on sustainable performance management in the service sector in ASEAN,” *Decision Science Letters*, vol. 14, no. 3, pp. 1–10, 2025, doi: 10.5267/j.dsl.2025.7.003.

[38] I. Chkhaidze, "Reasons, challenges and strategies for implementing agile in today's unpredictable environment," *Innovative Economics and Management*, vol. 10, no. 2, pp. 72–86, 2023, doi: 10.46361/2449-2604.10.2.2023.72-86.

[39] A. O. Temitope, "Agile and organizational culture: Fostering agile values and mindset," *International Journal of Scientific Research Archive*, vol. 7, no. 2, pp. 672-681, 2022, doi: 10.30574/ijsra.2022.7.2.0265.