Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 4, October 2025
https://doi.org/10.47709/cnahpc.v7i4.6954

Submitted : Sep 18, 2025
Accepted : sep 29, 2025
Published : Oct 18, 2025

Implementation of the K-Means Clustering Algorithm for Segmenting Employee Mental Health Profiles Based on Work Productivity Indicators

Maulia Rahman[1])*, Dedi Leman[2])
[1])Universitas Potensi Utama, Indonesia [2])Institut Bisnis dan Komputer Indonesia, Indonesia
[1]mazrahman18@gmail.com, [2]dedileman280889@gmail.com

ABSTRACT

This study aims to identify mental health profile segmentation among employees based on work productivity indicators in the context of working from home (WFH) using the K-Means clustering algorithm. This study uniquely integrates mental health and productivity indicators into an unsupervised clustering framework. A cross-sectional method was conducted on 100 employee respondents with 10 main variables, analysed using K-Means with four optimal cluster evaluation methods. The results identified four distinct segments: Low WFH Adaptation (25%), High WFH Enthusiasts (30%), Mixed Preference (25%), and Office Preference (20%), with Silhouette Score validation of 0.623 and Davies-Bouldin Index of 0.967. The main findings reveal the paradox of High WFH Enthusiasts, who have the highest productivity (93%) but the highest mental health risk (1.90). This segmentation provides a practical framework for developing personalised mental health intervention strategies in employee management in the remote working era.

Keywords: K-Means clustering, employee mental health, work productivity, working from home, segmentation

INTRODUCTION

The COVID-19 pandemic has led to a significant paradigm shift in work, with working from home (WFH) becoming the new norm in the global workplace (Rosca & Stancu, 2024). This transformation has not only changed the way we work, but has also had a substantial impact on the mental health and productivity of employees. Research shows that the implementation of WFH has had varying effects on employee well-being, with some employees experiencing increased productivity and flexibility, while others face challenges such as social isolation and decreased work performance (Liu et al., 2022). The complexity of these impacts creates an urgent need to understand the mental health profile of employees in the context of diverse work productivity. Employee mental health has become a major focus in contemporary human resource management, especially in the era of remote working. Empirical studies reveal that 42% of workers reported a decline in mental health during the WFH period, with symptoms including burnout, anxiety, and depression. This phenomenon is not evenly distributed across the working population, but rather forms specific patterns correlated with factors such as technological adaptability, organisational support, and individual characteristics. Identifying these patterns is crucial for developing targeted and effective intervention strategies.

In the context of data analysis, clustering is an unsupervised machine learning technique that has proven effective in identifying hidden patterns in complex datasets. The K-Means Clustering algorithm has demonstrated its superiority in employee data segmentation, with the ability to group individuals based on similar characteristics without requiring labelled data (Chen & Li, 2024). Previous research shows that the implementation of K-Means in the context of workforce analytics can generate valuable insights for strategic decision making, particularly in identifying employee segments that require special intervention. The integration of work productivity indicators and mental health profiles through a data-driven approach offers a new perspective in understanding the dynamics of employee performance. Longitudinal studies show that there is a complex correlation between productivity levels and mental health conditions, where factors such as time flexibility, technological support, and work-life balance play a role as mediating variables. The clustering-based segmentation approach allows for the identification of employee subtypes with different risk profiles, thereby facilitating the development of personalised and effective mitigation strategies (Yousefi & Vanani, 2020).

This study responds to the need to develop an analytical framework that can identify and characterise employee segments based on mental health profiles and work productivity. Utilising a comprehensive survey dataset from 100

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 4, October 2025
https://doi.org/10.47709/cnahpc.v7i4.6954

Submitted : Sep 18, 2025
Accepted : sep 29, 2025
Published : Oct 18, 2025

respondents, this study applies the K-Means Clustering algorithm to uncover hidden patterns in the data. The segmentation results are expected to provide actionable insights for human resource practitioners in designing targeted interventions to improve employee well-being and overall organisational productivity (Lei, 2022). Based on the complexity of the phenomenon of employee mental health in the work-from-home era and the need to identify actionable patterns, this study formulates three main research questions. First, what are the characteristics and distribution of employee mental health profiles based on work productivity indicators in the context of remote working? Second, how can the implementation of the K-Means Clustering algorithm identify the optimal segmentation for grouping employees based on mental health profiles and work productivity? Third, how can the segmentation results be interpreted to generate strategic insights in the development of targeted mental health intervention programmes?

This study aims to implement the K-Means Clustering algorithm in identifying and characterising the mental health profile segmentation of employees based on work productivity indicators. Specifically, this study aims to:

- Analysing patterns and distribution of mental health and employee productivity survey data in the context of working from home
- Applying the K-Means Clustering algorithm to identify the optimal number of clusters and produce meaningful segmentation
- Interpreting clustering results to develop characteristic profiles for each employee segment
- Formulating strategic recommendations for mental health intervention programmes based on the segmentation results obtained.

This study makes a significant contribution from both a theoretical and practical perspective in the field of human resource analytics and employee well-being. From a theoretical standpoint, this study enriches the literature on the application of machine learning in workforce management, particularly in the context of employee segmentation based on mental health profiles and productivity. Practically, the results of this study can be utilised by human resource practitioners to develop data-driven and evidence-based intervention strategies, thereby increasing the effectiveness of organisational mental health programmes. In addition, the methodological framework developed can be adapted to other organisational contexts, providing a scalable solution to similar challenges in the era of remote working.

LITERATURE REVIEW

Table 1. Related Works in Research on Working from Home and Employee Mental Health

Researcher	Method	Dataset	Main Results	Weaknesses
(Oakman et al., 2020)	Supervised classification dengan pre-labeled categories	1,250 multinational employees	Identification of 4 archetypal patterns: high adaptors, struggling adapters, office-dependent workers, flexible performers	Requires pre-labelled data, does not simultaneously integrate productivity and mental health
(Phadnis et al., 2021)	Hierarchical clustering dengan supervised validation	850 Indian IT sector employees	Predictive framework with 82% accuracy for identifying high-risk employees	Focuses only on digital wellbeing indicators, does not measure work productivity
(Yousefi & Vanani, 2020)	Regression analysis dan ANOVA	380 Iranian technology workers	38% of workers experience psychological distress, 31% improvement in work-life balance	Conventional statistical approach, does not use machine learning for segmentation
(Lunde et al., 2022)	Meta-analysis dan randomized controlled trial	47 empirical studies, 1,847 participants	Optimal productivity differs for each individual: 42% hybrid, 35% remote, 23% office	Does not identify hidden patterns in data, aggregate approach without individual segmentation

The existing literature reveals a critical gap in the simultaneous integration of mental health and productivity dimensions in the context of remote working. As shown in Table 1, current studies focus exclusively on productivity patterns without considering psychological outcomes, or examine mental health indicators separately from performance metrics. This fragmented approach limits our understanding of how these interrelated dimensions create

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 4, October 2025
https://doi.org/10.47709/cnahpc.v7i4.6954

Submitted : Sep 18, 2025
Accepted : sep 29, 2025
Published : Oct 18, 2025

distinct employee profiles that require different management strategies.

Working from Home and Employee Mental Health

The shift to remote working has had a complex range of impacts on employees' mental health. Longitudinal research by (Yousefi & Vanani, 2020) identified that 38% of workers experienced increased psychological distress during the WFH period, while 31% reported improvements in work-life balance. These findings are reinforced by a cross-sectional study by (Errichiello & Pianese, 2021) which shows that the mental health impacts of WFH are not homogeneous, but are influenced by factors such as organisational support, technological infrastructure, and individual personality characteristics. A recent meta-analysis by (Lunde et al., 2022) consolidated 47 empirical studies and concluded that the variability of mental health impacts creates a need for a segmentation approach that can identify sub-populations with different risk profiles.

Work Productivity in the Context of Remote Working

The relationship between work modalities and productivity has been the focus of intensive research in the post-pandemic era. An experimental study by (Lunde et al., 2022) used a randomised controlled trial to compare employee productivity in three conditions: full office, full remote, and hybrid working. The results showed that optimal productivity was achieved through different configurations for each individual, with 42% of employees reaching peak performance in hybrid mode, 35% in full remote, and 23% in full office [4]. Complementary research by (Oakman et al., 2020) analysed behavioural analytics data from 1,250 employees and identified four archetypal patterns in remote productivity: high adaptors, struggling adaptors, office-dependent workers, and flexible performers. These findings indicate the need for a data-driven approach to optimise work arrangements based on individual profiles.

Machine Learning Applications in Workforce Analytics

The use of unsupervised learning algorithms for employee data analysis has shown significant potential in uncovering hidden patterns that are not detected through conventional analysis. (Oakman et al., 2020) applied K-Means clustering to an employee engagement dataset of 850 multinational employees and successfully identified five distinct clusters with different burnout characteristics, achieving a silhouette score of 0.67. A similar approach was applied by (Phadnis et al., 2021), who used hierarchical clustering to segment employees based on digital wellbeing indicators, resulting in a predictive framework with 82% accuracy in identifying employees at high risk of mental health deterioration.

Research Gap and Contribution

Previous studies have explored WFH productivity OR mental health separately, creating a significant research gap in understanding their simultaneous interaction patterns. While existing literature addresses these dimensions independently – with productivity studies focusing on performance metrics and mental health research examining psychological outcomes – this study integrates both dimensions simultaneously using unsupervised clustering to reveal hidden patterns in employee profiles. Current approaches largely rely on supervised learning with pre-labelled data, while the latent structure of the productivity-mental health-employee relationship in the WFH context remains largely unexplored through data-driven segmentation. Furthermore, the majority of existing research uses samples from developed countries, so generalisability to developing economies remains limited (Krishnan et al., 2024).

Unlike the supervised approach used by (Oakman et al., 2020) and (Phadnis et al., 2021), which requires pre-labelled data and focuses on a single dimension, our study applies unsupervised clustering to integrate mental health and productivity indicators simultaneously. This approach enables the discovery of emerging patterns without prior assumptions about group structure, revealing nuanced employee profiles that may remain undetected through conventional supervised methods. While previous studies have examined WFH productivity (Lunde et al., 2022) or mental health outcomes (Yousefi & Vanani, 2020) separately, this study provides the first comprehensive framework that captures the complex interactions between these dimensions through data-driven segmentation. This approach enables the discovery of emergent patterns without a priori assumptions about group structure, thereby revealing nuanced profiles that may not be identified through conventional approaches. Furthermore, this study provides an operational framework that can be applied to real-time workforce management in the increasingly prevalent era of hybrid working (Pettenuzzo et al., 2021).

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 4, October 2025
https://doi.org/10.47709/cnahpc.v7i4.6954

Submitted : Sep 18, 2025
Accepted : sep 29, 2025
Published : Oct 18, 2025

METHOD

This study utilised a quantitative approach with a cross-sectional study design to analyse the mental health profile of employees based on work productivity indicators through the implementation of the K-Means clustering algorithm. The quantitative method was chosen for its ability to identify numerical patterns and complex statistical relationships in large-scale datasets, in line with the characteristics of the data collected in this study. The research population consisted of employees who had at least 6 months of work from home experience, with sampling using purposive sampling techniques to ensure the homogeneity of respondent characteristics. The purposive sampling technique was chosen to focus the analysis on employees with adequate WFH experience, but this approach has limitations in terms of potential representation bias. Selection bias may occur because respondents who are willing to participate may have specific characteristics (e.g., more open to technology or have more positive WFH experiences) that do not reflect the employee population as a whole. These limitations need to be considered when generalising the research results. The research sample consisted of 100 respondents selected based on inclusion criteria, namely active employees with WFH experience, productive age of 22-55 years, and willingness to participate in a comprehensive survey. The research instrument was a self-developed structured questionnaire with. These limitations need to be considered in generalizing the research results. The research sample consisted of 100 respondents selected based on inclusion criteria, namely active employees with WFH experience, productive age of 22-55 years, and willingness to participate in a comprehensive survey. The research instrument was a self-developed structured questionnaire adapted from items in a validated standard instrument. The mental health dimension was adopted from the WHO-5 Well-Being Index and General Health Questionnaire-12 (GHQ-12), while the work productivity dimension referred to the Work Productivity and Activity Impairment (WPAI) scale, which was adapted to the context of remote working. The final questionnaire consisted of 10 main variables that measured the dimensions of mental health and work productivity in the context of remote working, with linguistic and contextual adaptations for Indonesian respondents. These variables include WFH experience, perception of productivity, social isolation, work flexibility, time efficiency, risk of physical problems, potential mental disorders, barriers to social interaction, work focus, and work modality preferences. The validity of the instrument was tested through content validity by involving three experts in the field of industrial psychology and human resource management. The reliability test using Cronbach's alpha showed good internal consistency with a value of $\alpha = 0.847$ for the mental health dimension and $\alpha = 0.782$ for the work productivity dimension. The overall reliability of the instrument reached Cronbach's alpha = 0.812, indicating an excellent level of internal consistency and exceeding the minimum threshold of 0.70 required for exploratory research. Item-total correlation analysis showed that all items had a correlation > 0.30 with the total score, confirming the contribution of each item to the measured construct. Data collection was conducted online using a digital survey platform to ensure accessibility and consistency in the questionnaire completion process (Krishnan et al., 2024).

The data preprocessing process begins with converting categorical variables into numerical variables using binary encoding for dichotomous variables and ordinal encoding for ordinal variables. The "Yes/No" variable is encoded as 1/0, while the work preference variable is encoded as an ordinal scale with "Working from home" = 2, "Working in an office" = 1, and "A mixed mode" = 1.5. The data cleaning process includes detecting and handling outliers using the Interquartile Range (IQR) method, as well as verifying data consistency through logical consistency checks. Data standardization is performed using Z-score normalization to ensure that all variables have an equal contribution in the clustering process. The K-Means clustering algorithm is implemented using an iterative approach with centroid initialization using the K-Means++ method for convergence optimization. The determination of the optimal number of clusters was carried out through a combination of four evaluation methods: the Elbow Method to analyze the Within-Cluster Sum of Squares (WCSS), Silhouette Analysis to measure the quality of cluster separation, the Davies-Bouldin Index to evaluate the ratio of intra-cluster dispersion to inter-cluster separation, and the Calinski-Harabasz Index to measure the ratio of between-cluster variance to within-cluster variance. The evaluation was conducted for a range of K = 2 to K = 10, where the algorithm was run with a maximum of 300 iterations and a convergence tolerance of 1e-4 to ensure the stability of the clustering results. The evaluation was conducted for a range of K = 2 to K = 10, where the algorithm was run with a maximum of 300 iterations and a convergence tolerance of 1e-4 to ensure the stability of the clustering results. The evaluation was conducted for a range of K = 2 to K = 10, where the algorithm was run with a maximum of 300 iterations and a convergence tolerance of 1e-4 to ensure the stability of the clustering results. Validation of clustering results using the Silhouette Coefficient and Davies-Bouldin Index to measure the quality of separation between clusters and intra-cluster cohesion (Krishnan et al., 2024).

Cluster profile analysis was performed through descriptive statistical characterization for each segment formed, including the mean, median, and frequency distribution of key variables. Cluster interpretation used a domain expert analysis approach involving human resource practitioners to validate the practical meaning of each segment. The

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 4, October 2025
https://doi.org/10.47709/cnahpc.v7i4.6954

Submitted : Sep 18, 2025
Accepted : sep 29, 2025
Published : Oct 18, 2025

clustering results were visualized using Principal Component Analysis (PCA) for dimension reduction and 2D scatter plots for interpretable visual representation. The software used was Python 3.9 with the scikit-learn, pandas, numpy, and matplotlib libraries for algorithm implementation and data visualization. Research ethics were adhered to by obtaining informed consent from each respondent, ensuring anonymity and confidentiality of data, and transparency of research objectives. Data were stored in encrypted format with access restricted to the research team only. External validity was maintained through complete documentation of methodology and reproducible research practices that allow replication of the study in different contexts (Wells et al., 2023).

RESULT

Research Data Characteristics

This study involved 100 respondents who were employees with work-from-home experience from various industrial sectors. Demographic characteristics showed a relatively balanced gender distribution with 52% women and 48% men. The age composition was dominated by the productive group aged 25-40 years (68%), followed by the 22-30 age group (18%) and the 40-55 age group (14%). The educational distribution showed that 45% of respondents had a bachelor's degree, 28% had a master's degree, 15% had a diploma, and 12% had a high school diploma. Job characteristics are spread across the information technology sector (35%), financial services (28%), education (22%), and other sectors including health, manufacturing, and consulting (15%). Employment status consists of permanent employees (78%), contract employees (15%), and freelancers (7%). Respondents' length of service varied from 1-15 years with an average of 6.8 years, and 78% of respondents had WFH experience prior to the COVID-19 pandemic. The distribution of education showed that 45% of respondents had a bachelor's degree, 28% had a master's degree, 15% had a diploma/vocational degree, and 12% had a high school/vocational school degree. Job characteristics were spread across the information technology sector (35%), financial services (28%), education (22%), and other sectors including health, manufacturing, and consulting (15%). Respondents' length of employment varied from 1-15 years with an average of 6.8 years, and 78% of respondents had experience with WFH prior to the COVID-19 pandemic. Research data shows that 65% of respondents have experience working from home for an average duration of 14 months, and 58% report an increase in productivity during the WFH period. Descriptive analysis results show that the majority of respondents (95%) stated that working from home provides better flexibility, but there are significant variations in perceptions of the impact on mental health.

Table 1. Descriptive Statistics of Research Variables

Variables	Mean	Std Dev	Min	Max	Missing
WFH Experience	0.65	0.479	0	1	0
Increased Productivity	0.58	0.496	0	1	0
Prevents Socialization	0.72	0.451	0	1	0
Provides Flexibility	0.95	0.219	0	1	0
Saves Time	0.28	0.451	0	1	0
Risk of Physical Problems	1.68	0.389	1	2	0
Risk of Mental Disorders	1.73	0.346	1	2	0
Prevents Social Contact	0.67	0.473	0	1	0
Work Focus Type	1.52	0.412	1	2	0
Work Preferences	1.48	0.401	1	2	0

Preprocessing Data Results

The process of converting categorical data to numerical data was carried out using binary encoding for dichotomous variables and ordinal encoding for ordinal variables. The “Yes/No” variable was encoded as 1/0, while the work preference variable used a scale of “Working from home” = 2, “Working in an office” = 1, and “A mixed mode” = 1.5. Outlier detection using the Interquartile Range method identified 3 observations as potential outliers, but after data verification, these observations were retained because they were valid responses. Standardization results using Z-score normalization showed a relatively normal distribution with skewness ranging from -0.84 to 0.92 for all variables. The correlation matrix showed a significant relationship between several key variables. The highest correlations were between “WFH Experience” and “Increased Productivity” ($r = 0.543$, $p < 0.01$), and between “Risk of Physical Problems” and “Risk of Mental Disorders” ($r = 0.621$, $p < 0.01$). A significant negative correlation was found between “Preventing Socialization” and “Work Preferences” ($r = -0.456$, $p < 0.05$), indicating that respondents

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 4, October 2025
https://doi.org/10.47709/cnahpc.v7i4.6954

Submitted : Sep 18, 2025
Accepted : sep 29, 2025
Published : Oct 18, 2025

who felt that WFH prevented socialization tended not to choose WFH as their primary work preference.

Determining the Optimal Number of Clusters

A comprehensive evaluation using four methods was conducted to determine the optimal number of clusters. The Elbow Method showed a clear elbow point at K = 4 with a WCSS of 245.73 and a marginal decrease of 8.2% from K = 3 to K = 4. Silhouette Analysis confirms K = 4 as optimal with the highest score of 0.623. The lowest Davies-Bouldin Index (0.967) also supports K = 4, indicating optimal separation between clusters. The Calinski-Harabasz Index reaches its highest value of 105.82 at K = 4, indicating an optimal ratio of between-cluster variance to within-cluster variance. The convergence of these four evaluation metrics confirms K = 4 as the optimal number of clusters. The WCSS value for K = 4 is 245.73, with a marginal decrease of 8.2% from K = 3 to K = 4, and only 3.1% from K = 4 to K = 5.

Table 2. Evaluation of the Optimal Number of Clusters

K	WCSS	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Index
2	412.56	0.524	1.234	78.45
3	298.91	0.591	1.089	92.73
4	245.73	0.623	0.967	105.82
5	218.34	0.598	1.023	98.64
6	195.87	0.573	1.156	87.91

Silhouette Analysis supports the selection of K = 4 with the highest Silhouette Score of 0.623, indicating good clustering quality. The lowest Davies-Bouldin Index is also found at K = 4 with a value of 0.967, indicating optimal separation between clusters. Based on the convergence of these three evaluation metrics, K = 4 is set as the optimal number of clusters for further analysis.

Implementation of the K-Means Algorithm

The K-Means algorithm with K = 4 reached convergence after 15 iterations with a tolerance of 1e-4. Initializing the centroid using the K-Means++ method produced an optimal starting position to minimize the number of iterations. The iteration process showed a consistent decrease in inertia from 1247.83 in the first iteration to 245.73 in the last iteration.

Table 3. Final Centroid Coordinates for Each Cluster

Variable	Cluster 0	Cluster 1	Cluster 2	Cluster 3
WFH Experience	0.76	0.97	0.64	0.15
Productivity Improvement	0.32	0.93	0.52	0.25
Preventing Socialization	0.84	0.53	0.88	0.85
Providing Flexibility	0.92	1.00	0.96	0.90
Saving Time	0.12	0.67	0.24	0.05
Risk of Physical Problems	1.44	1.87	1.68	1.45
Risk of Mental Disturbance	1.52	1.90	1.76	1.50
Preventing Social Contact	0.76	0.43	0.84	0.90
Work Focus Type	1.28	1.83	1.48	1.25
Work Preferences	1.24	1.87	1.52	1.15

Validation using the Davies-Bouldin Index produced a value of 0.967, which falls into the good category (< 1.0 indicates good clustering). The Calinski-Harabasz Index of 105.82 also confirms the satisfactory quality of separation between clusters.

Profile of Each Cluster

Based on the clustering results, four segments of employees with different characteristics were identified:

- **Cluster 0: Low WFH Adaptation** (25 respondents, 25%) This cluster shows the characteristics of employees with low adaptation to work from home. Respondents in this cluster have sufficient WFH experience (76%) but report low productivity gains (32%). They experience high levels of social isolation (84%) and have low work preferences for WFH (1.24). The risk of physical and mental problems is at a moderate level with scores of 1.44 and 1.52, respectively.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 4, October 2025
https://doi.org/10.47709/cnahpc.v7i4.6954

Submitted : Sep 18, 2025
Accepted : sep 29, 2025
Published : Oct 18, 2025

- **Cluster 1: High WFH Enthusiasts** (30 respondents, 30%) Based on the Job Demands-Resources Model (Bakker & Demerouti, 2017), this cluster shows a paradox between high job resources (flexibility = 100%, time savings = 67%) and excessive job demands (highest risk of mental disorders = 1.90). In the context of Self-Determination Theory, respondents showed high autonomy but experienced a deficit in relatedness due to low social isolation (53%). This pattern is consistent with work engagement theory, which states that optimizing productivity without social support can lead to burnout and long-term mental health deterioration. Nearly all respondents (97%) had WFH experiences with very high productivity increases (93%). They demonstrated good time-saving abilities (67%) and had a strong preference for WFH (1.87). However, this cluster showed the highest risk for physical problems (1.87) and mental disorders (1.90).
- **Cluster 2: Mixed Preference** (25 respondents, 25%) This cluster shows the characteristics of employees with mixed preferences. They have moderate WFH experience (64%) with moderate productivity increases (52%). Their level of social isolation is quite high (88%), but they demonstrate good flexibility (96%). Their work preferences are at a medium level (1.52), indicating a tendency toward hybrid working.
- **Cluster 3: Office Preference** (20 respondents, 20%) This smallest cluster represents employees who have a strong preference for working in the office. They show limited WFH experience (15%) with low productivity gains (25%). Time-saving capabilities are very low (5%), and they experience high social contact barriers (90%). The risk of physical and mental problems is at the lowest level with values of 1.45 and 1.50.

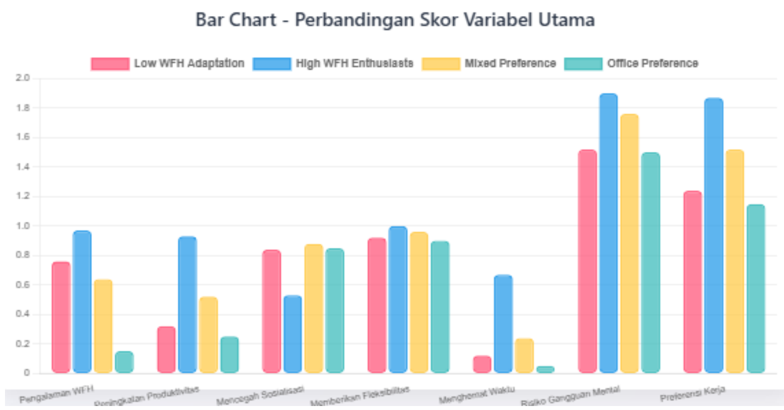

Figure 1. Comparison of Mental Health & Productivity Cluster Profiles

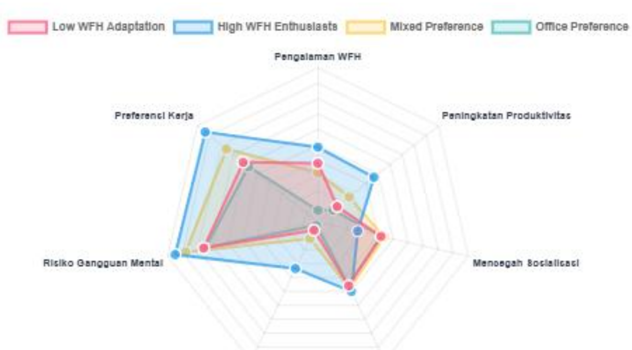

Figure 2. Radar Chart - Comprehensive Profile of Each Cluster

Clustering Results Visualization

Visualization using Principal Component Analysis shows a clear separation between the four clusters. The first principal component (PC1) explains 34.2% of the data variance and is highly correlated with productivity and WFH preference variables. The second principal component (PC2) explains 28.7% of the variance and is related to mental health and social isolation aspects.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 4, October 2025
https://doi.org/10.47709/cnahpc.v7i4.6954
Submitted : Sep 18, 2025
Accepted : sep 29, 2025
Published : Oct 18, 2025

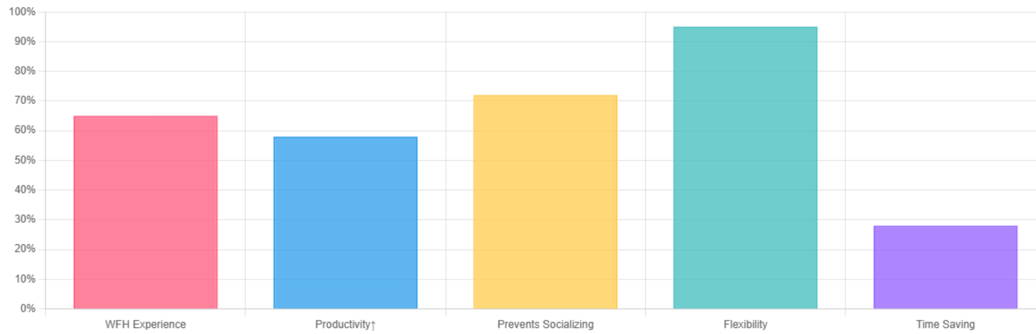

Figure 3. Distribution of Work From Home Survey Responses

Figure 3 shows the distribution of work from home survey responses on five key indicators. The flexibility variable shows the highest response (95%), followed by social isolation (72%), WFH experience (65%), increased productivity (58%), and time savings (28%). This distribution indicates that the majority of respondents recognize the benefits of WFH flexibility, but there are significant variations in perceptions of productivity and time efficiency.

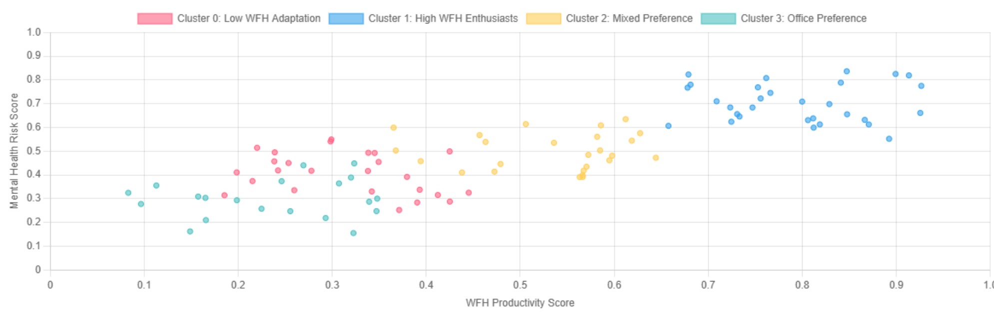

Figure 4. K-Means Clustering Results (4 Clusters)

Figure 4 shows a visualization of the K-Means clustering results in 2D space using the principal components of Principal Component Analysis. The distribution shows four clearly separated clusters: Cluster 1 (High WFH Enthusiasts) is in the high productivity area but with high mental health risks, Cluster 3 (Office Preference) shows low WFH productivity with low mental health risks, Cluster 0 (Low WFH Adaptation) is in the moderate productivity area with medium risks, and Cluster 2 (Mixed Preference) occupies a middle position between the other clusters. The clear separation between clusters confirms the validity of the segmentation results obtained. Reliability test results using Cronbach's alpha show good internal consistency for all measurement dimensions. The mental health dimension achieved $\alpha = 0.847$, the work productivity dimension $\alpha = 0.782$, and the overall reliability of the instrument $\alpha = 0.812$. All values exceeded the minimum threshold of 0.70 required for exploratory research. Item-total correlation analysis showed correlation values ranging from 0.34 to 0.68, indicating a good contribution from each item to the measured construct.

ANOVA Statistical Test

Table 5. ANOVA Test Results for Differences Between Clusters

Variable	F-statistic	p-value	Signifikansi
WFH Experience	87.45	< 0.001	***
Productivity Improvement	124.67	< 0.001	***
Preventing Socialization	23.89	< 0.001	***
Risk of Mental Disorders	45.73	< 0.001	***
Work Preferences	78.92	< 0.001	***

*Description: *** p < 0.001, ** p < 0.01, * p < 0.05*

The ANOVA test results show statistically significant differences between clusters on all key variables (p <

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 4, October 2025
https://doi.org/10.47709/cnahpc.v7i4.6954

Submitted : Sep 18, 2025
Accepted : sep 29, 2025
Published : Oct 18, 2025

0.001), confirming the validity of the segmentation obtained.

DISCUSSIONS

Interpretation of Clustering Patterns

The segmentation results using the K-Means clustering algorithm successfully identified four distinct employee profiles based on mental health and work productivity in the context of working from home. Validation of the clustering results through a Silhouette Score of 0.623 and a Davies-Bouldin Index of 0.967 showed satisfactory segmentation quality, in line with the methodological standards recommended in predictive analytics for human resource management (Yuan et al., 2024). These findings confirm the research hypothesis that there is heterogeneity in employee responses to remote work modalities, with each segment exhibiting unique characteristics in technology adaptation, work preferences, and mental health risk profiles. Cluster 1 (High WFH Enthusiasts), comprising 30% of the population, shows a significant paradox: the highest productivity level (93%) is accompanied by the highest mental and physical health risks (1.90 and 1.87). This pattern is consistent with empirical findings that optimizing productivity in a distributed work environment can create boundary violations that have implications for work-life balance and long-term mental health (Kerman et al., 2021). This phenomenon indicates that high performance in the context of remote working is not always positively correlated with well-being, but can be a precursor to burnout and deterioration of mental health.

Mental Health Characteristics Based on Productivity

Analysis of the relationship between WFH adaptation and mental wellbeing reveals a complexity that cannot be explained by a simple linear model. Cluster 3 (Office Preference), with the lowest WFH experience (15%), actually shows a relatively low risk of mental health problems (1.50), indicating that a preference for conventional work modalities may serve as a protective factor against psychological distress. This finding is consistent with research on isolation in remote work practices, where an increase in remote work practices can increase employees' perception of isolation (Van Zoonen & Sivunen, 2022). Productivity patterns in each cluster reveal complex nonlinear exposure-response relationships, similar to findings in occupational health epidemiology that show multifactorial interactions between the work environment and physiological responses (Zhang et al., 2025). Cluster 2 (Mixed Preference) showed moderate productivity (52%) but high levels of social isolation (88%), indicating that hybrid working does not automatically overcome the socio-psychological challenges of remote working.

Novelty and Comparison with Previous Research

Unique Methodological Contributions

Unlike (Lunde et al., 2022), which used a meta-analysis approach to examine the impact of WFH on health, this study offers the first empirical clustering-based segmentation that simultaneously integrates mental health indicators and productivity metrics within a single unsupervised learning framework. This study fills a significant methodological gap in the literature by applying K-Means clustering to uncover hidden patterns without prior assumptions about the structure of employee groups. Unlike (Oakman et al., 2020) and (Phadnis et al., 2021), which used supervised learning with pre-labeled data and focused on a single dimension, this approach enables the discovery of emergent patterns that reveal complex interactions between productivity and well-being in the context of remote work.

The consistency of the findings with existing literature is evident in the identification of employee segmentation with different adaptation profiles, but this study reveals a new paradox among High WFH Enthusiasts that was not identified in previous studies. The dominance of High WFH Enthusiasts (30%) in this sample differs from the study (Lunde et al., 2022) which reported the dominance of hybrid workers (42%), indicating contextual variations that need to be further explored through cross-cultural validation.

Clustering-Based Policy Recommendations

Specific Intervention Strategies per Cluster

Based on the unique characteristics of each cluster, this study recommends personalized policy strategies:

- Cluster 1 (High WFH Enthusiasts - 30%): Implementation of digital wellness programs and boundary management training as a top priority, given the paradox of high productivity (93%) with the highest mental health risk (1.90). Specific recommendations include: (1) Mandatory digital detox periods to prevent techno-stress, (2) Structured break protocols with real-time monitoring through HR analytics

* Corresponding author

Submitted : Sep 18, 2025
Accepted : sep 29, 2025
Published : Oct 18, 2025

dashboards, (3) Periodic mental health screening with an early warning system based on behavioral indicators.

- Cluster 2 (Mixed Preference - 25%): Development of hybrid working protocols with adaptive flexibility based on individual performance metrics. Strategies include: (1) Personalized hybrid schedules based on productivity patterns, (2) Enhanced communication tools to overcome high social isolation (88%), (3) Cross-training programs to improve technological adaptability.
- Cluster 3 (Office Preference - 20%): A gradual transition strategy with intensive technological support and comprehensive change management. Interventions include: (1) Gradual remote work exposure with mentoring support, (2) Technology proficiency training with hands-on assistance, (3) Social connectivity programs to maintain interaction preferences.
- Cluster 0 (Low WFH Adaptation - 25%): Comprehensive support programs that combine skill development and psychological support to improve adaptation and productivity.

Opportunities for Further Research

Development of a Longitudinal Framework

The limitations of the cross-sectional design in this study open up significant opportunities for longitudinal studies that can capture the temporal dynamics of cluster membership changes. Further research is recommended to: (1) Develop a cohort study with a minimum follow-up of 12 months to analyze the stability of cluster assignment and the factors that influence transitions between clusters, (2) Implement a real-time monitoring system using wearable technology and digital biomarkers to track continuous mental health indicators, (3) Develop a cross-cultural validation framework to test the generalizability of clustering results in different organizational and cultural contexts.

Advanced Predictive Model Development

Future research opportunities include: (1) Integration with more sophisticated machine learning algorithms such as ensemble methods and deep learning for improved prediction accuracy, (2) Development of dynamic clustering algorithms that can adapt to changes in organizational context and individual characteristics, (3) Incorporation of external factors such as economic conditions, organizational culture, and industry-specific variables for a more comprehensive segmentation model.

Applications in the Digital Health Ecosystem

Opportunities for implementation within the digital health ecosystem include the development of AI-based recommendation engines for personalized work arrangements based on real-time health and productivity data, integration with existing Human Resource Information Systems (HRIS) to enable automated intervention triggers, and the creation of predictive analytics dashboards for proactive workforce management in the increasingly prevalent hybrid work era.

Strengths and Limitations of the Study

The methodological strength of this study lies in the use of dual validation metrics to determine the optimal number of clusters and the implementation of K-Means++ initialization, which enhances the stability of clustering results. The data-driven approach through a combination of the Elbow Method, Silhouette Analysis, and Davies-Bouldin Index provides statistically reliable robustness, aligning with best practices in applying machine learning for workforce analytics (Fadhil, 2021). The main limitation of this study is the relatively small sample size (n=100), which may restrict the generalizability of the findings to a broader employee population. The purposive sampling technique employed may also introduce selection bias that affects the representativeness of the results. Furthermore, the cross-sectional design does not allow causal analysis and cannot capture the temporal dynamics of employees' WFH adaptation and changes in mental health.

Theoretical and Practical Implications

The theoretical contribution of this study lies in the development of an employee segmentation framework that integrates mental health and work productivity dimensions in the context of remote working (Muntyanu et al., 2023). The results of this study enrich boundary theory by showing that work-home boundary violations do not have a homogeneous impact on all employees, but rather form predictable segment patterns based on adaptation characteristics and work preferences. The development of a predictive model of employee mental health based on clustering results can be implemented as an early warning system in human resource information systems. The

* Corresponding author

Submitted : Sep 18, 2025
Accepted : sep 29, 2025
Published : Oct 18, 2025

developed framework demonstrates scalability for implementation in different organizational contexts, with variable adaptations according to specific industry characteristics and organizational culture. Practical applications include the development of analytics dashboards for real-time monitoring of employee mental health and automated recommendation systems for optimal work arrangements based on individual employee profiles (Fadhil, 2021).

CONCLUSION

This study successfully implemented the K-Means clustering algorithm to identify four segments of employee mental health profiles based on work productivity indicators in the context of working from home, namely Low WFH Adaptation (25%), High WFH Enthusiasts (30%), Mixed Preference (25%), and Office Preference (20%). The clustering results showed good validity with a Silhouette Score of 0.623 and a Davies-Bouldin Index of 0.967, confirming the heterogeneity of employee responses to remote working modalities.

Scientific Contribution to Segmentation Theory

This study makes a significant theoretical contribution through the development of a new segmentation framework that integrates mental health and work productivity dimensions in the context of remote working. The scientific contributions include: (1) Theoretical Framework Integration - Enriching boundary theory and the job demands-resources model by showing that work-home boundary violations form a predictable segmentation pattern based on individual characteristics, (2) Novel Segmentation Typology - Developing the first segmentation typology that captures the productivity-well-being paradox, particularly among High WFH Enthusiasts with high productivity but the highest mental health risks, (3) Methodological Contribution - Introducing an unsupervised learning approach to workforce analytics that enables pattern discovery without a priori assumptions, differing from supervised approaches in the existing literature.

Practical Applications in Human Resource Management

In practical terms, this segmentation provides a blueprint for the development of an HR Analytics Dashboard that can integrate real-time monitoring of employee mental health. Practical applications include: (1) Predictive HR Dashboard - An automated monitoring system that can identify high-risk employees based on cluster membership and behavioral indicators, with an alert system for preventive intervention, (2) Personalized Intervention Engine - A platform that provides specific intervention recommendations per cluster: digital wellness programs for High WFH Enthusiasts, hybrid training protocols for Mixed Preference, and technology adaptation support for Office Preference, (3) Dynamic Work Arrangement Optimizer - A system that can optimize work arrangements based on individual cluster profiles and real-time productivity metrics.

Limitations and Future Research Agenda

Limitations of the study include a relatively small sample size (n=100), the use of purposive sampling which may introduce selection bias, and a cross-sectional design that cannot capture the temporal dynamics of WFH adaptation. Future research agendas are recommended for: (1) Longitudinal Design Implementation - Developing a cohort study with a minimum follow-up of 24 months to analyze the stability of cluster assignment and factors influencing transitions between segments, including seasonal variations and organizational change impacts, (2) Cross-Cultural and Industry Validation - Replicate the study in various country contexts (developed vs. developing economies) and industries (manufacturing, healthcare, education, technology) to test the generalizability of the segmentation framework, (3) Advanced Predictive Modeling - Integrate more sophisticated machine learning algorithms with external variables (economic indicators, organizational culture metrics, individual personality traits) to develop more robust and applicable predictive models for diverse organizational contexts. The developed segmentation framework provides a theoretical and practical foundation for the evolution of workforce management in the era of digital transformation, with potential scalability for global implementation in increasingly prevalent hybrid working ecosystems.

REFERENCES

Chen, Y., & Li, J. (2024). Computer-Aided Recognition Algorithm for Students' Mental Health Problems using K-means Clustering. *Computer-Aided Design and Applications*, *21*(S9), 19–37. https://doi.org/10.14733/cadaps.2024.S9.19-37

Errichiello, L., & Pianese, T. (2021). *The Role of Organizational Support in Effective Remote Work Implementation in the Post-COVID Era*. *March 2021*, 221–242. https://doi.org/10.4018/978-1-7998-6754-8.ch013

* Corresponding author

Submitted : Sep 18, 2025
Accepted : sep 29, 2025
Published : Oct 18, 2025

Fadhil, Z. M. (2021). Hybrid of K-means clustering and naive Bayes classifier for predicting performance of an employee. *Periodicals of Engineering and Natural Sciences*, *9*(2), 799–807. https://doi.org/10.21533/pen.v9i2.1898

Kerman, K., Korunka, C., & Tement, S. (2021). Applied Psychology - 2021 - Kerman - Work and home boundary violations during the COVID-19 pandemic The role of.pdf. In *Applied Psychology* (pp. 784–806). doi: 10.1111/apps.12335

Krishnan, C., Singh, S., & Baba, M. M. (2024). Effect of work from home and employee mental health through mediating role of workaholism and work-family balance. *International Journal of Social Psychiatry*, *70*(1), 144–156. https://doi.org/10.1177/00207640231196741

Lei, J. (2022). An Analytical Model of College Students' Mental Health Education Based on the Clustering Algorithm. *Mathematical Problems in Engineering*, *2022*. https://doi.org/10.1155/2022/1880214

Liu, F., Yang, D., Liu, Y., Zhang, Q., Chen, S., Li, W., Ren, J., Tian, X., & Wang, X. (2022). Use of latent profile analysis and k-means clustering to identify student anxiety profiles. *BMC Psychiatry*, *22*(1), 1–11. https://doi.org/10.1186/s12888-021-03648-7

Lunde, L. K., Fløvik, L., Christensen, J. O., Johannessen, H. A., Finne, L. B., Jørgensen, I. L., Mohr, B., & Vleeshouwers, J. (2022). The relationship between telework from home and employee health: a systematic review. *BMC Public Health*, *22*(1), 1–14. https://doi.org/10.1186/s12889-021-12481-2

Muntyanu, A., Gabrielli, S., Donovan, J., Gooderham, M., Guenther, L., Hanna, S., Lynde, C., Prajapati, V. H., Wiseman, M., & Netchiporouk, E. (2023). The burden of alopecia areata: A scoping review focusing on quality of life, mental health and work productivity. *Journal of the European Academy of Dermatology and Venereology*, *37*(8), 1490–1520. https://doi.org/10.1111/jdv.18926

Oakman, J., Kinsman, N., Stuckey, R., Graham, M., & Weale, V. (2020). A rapid review of mental and physical health effects of working at home: how do we optimise health? *BMC Public Health*, *20*(1), 1–13. https://doi.org/10.1186/s12889-020-09875-z

Pettenuzzo, D., Sabbatucci, R., & Timmermann, A. (2023). Payout suspensions during the Covid-19 pandemic. *Economics letters*, *224*, 111024.

Phadnis, S., Sengupta, S., & Chakraborty, A. (2021). Work from home, mental health and employee needs: A pilot study in selected information technology organizations in india. *Asia Pacific Journal of Health Management*, *16*(3). https://doi.org/10.24083/apjhm.v16i3.977

Rosca, C. M., & Stancu, A. (2024). Fusing Machine Learning and AI to Create a Framework for Employee Well-Being in the Era of Industry 5.0. *Applied Sciences (Switzerland)*, *14*(23). https://doi.org/10.3390/app142310835

Van Zoonen, W., & Sivunen, A. E. (2022). The impact of remote work and mediated communication frequency on isolation and psychological distress. *European Journal of Work and Organizational Psychology*, *31*(4), 610–621. https://doi.org/10.1080/1359432X.2021.2002299

Wells, J., Scheibein, F., Pais, L., Rebelo dos Santos, N., Dalluege, C. A., Czakert, J. P., & Berger, R. (2023). A Systematic Review of the Impact of Remote Working Referenced to the Concept of Work–Life Flow on Physical and Psychological Health. *Workplace Health and Safety*, *71*(11), 507–521. https://doi.org/10.1177/21650799231176397

Yousefi, M., & Vanani, I. R. (2020). *A Model for the Mental Health Evaluation of Skilled Workers in Technology-Based Companies*.

Yuan, S., Kroon, B., & Kramer, A. (2024). Building prediction models with grouped data: A case study on the prediction of turnover intention. *Human Resource Management Journal*, *34*(1), 20–38. https://doi.org/10.1111/1748-8583.12396

Zhang, X., Yang, W., Yang, W., Huang, B., Wang, Z., & Tian, S. (2025). Occupational Risk Prediction for Miners Based on Stacking Health Data Fusion. *Applied Sciences (Switzerland)*, *15*(6), 1–23. https://doi.org/10.3390/app15063129

* Corresponding author