ISSN 2722-2039

AI Adoption Patterns in Banking: A PCA-Based K-Means Clustering Analysis Using Evident AI Index Rankings

Catherine Ngo [a,1,*], Orson Chi [b]

[a] Department of Agriculture, Food, and Resource Sciences, University of Maryland Eastern Shore, Princess Anne, Maryland, United States
[b] Department of Computer Science and Engineering Technology, University of Maryland Eastern Shore, Princess Anne, Maryland, United States
[1] ctngo@umes.edu
* corresponding author

ARTICLE INFO

Article history
Received January 28, 2025
Revised March 24, 2025
Accepted June 24, 2025

Keywords
Banking Industry
AI Adoption
Evident AI Index
Principal Component Analysis
K-Means Clustering

ABSTRACT

The study investigated the adoption of AI in the banking sector using a PCA-based k-means clustering method, drawing on data from the Evident AI Index Rankings. The objective was to identify distinct patterns in banks' integration and use of AI technologies, with an emphasis on talent, innovation, leadership, and transparency. Utilizing PCA for dimensionality reduction, the study distilled the intricate aspects of AI adoption into fundamental components, thereby improving the comprehension of clustering patterns among banks. The k-means clustering identified unique segments within the sector, such as early AI adopters, innovation leaders, and conservative implementers, each exhibiting distinct levels of AI maturity and application focus. These findings provided valuable insights into the competitive landscape of AI utilization in banking, highlighting leading institutions in AI-driven transformation and those encountering adoption challenges. The insights from this analysis offered practical implications for stakeholders, guiding strategies for improved AI integration and competitive positioning. The study emphasizes the significance of data-driven benchmarking tools, such as the Evident AI Index, in assessing and guiding technological evolution across the sector.

1. Introduction

The rapid advancement of artificial intelligence (AI) has significantly reshaped operational and strategic paradigms within the banking sector. As financial institutions navigate the complexities of digital transformation, the adoption of AI technologies has emerged as a key differentiator for achieving competitive advantage, enhancing service delivery, and fostering innovation [1], [2]. However, the pace and scope of AI integration vary across the industry, influenced by organizational readiness, leadership vision, technological capacity, and regulatory environments [3]. In this context, understanding how different banks approach and implement AI has become critical for stakeholders seeking to assess competitive positioning and formulate data-driven strategies. This study investigates the patterns of AI adoption in the banking industry through a data-centric lens, employing a principal component analysis (PCA)-based k-means clustering methodology. Drawing on the Evident AI Index Rankings—a comprehensive benchmarking tool that evaluates banks across dimensions such as talent, innovation, leadership, and transparency—this analysis simplifies complex AI adoption metrics into interpretable groupings [4]. The application of PCA effectively reduces the dimensionality of the dataset, enabling the capture of the most significant variances in the data and highlighting structural patterns in AI engagement among banks [5].

The clustering analysis reveals distinct AI adoption profiles within the sector. Notably, the study identifies groups such as early AI adopters, characterized by robust innovation capacity and strategic leadership in AI integration; innovation leaders, who drive AI research and development yet may encounter implementation bottlenecks; and conservative implementers, who are slower to adopt advanced AI tools [6]. These clusters reflect varying levels of AI maturity and deployment focus, offering a nuanced understanding of how banks align their digital strategies with evolving technological opportunities and constraints. By mapping these adoption patterns, the study provides critical insights into the competitive landscape of AI utilization in banking. The findings highlight institutions at the forefront of AI-driven transformation and reveal the structural and strategic challenges that others face. Moreover, clustering techniques enhance strategic clarity, enabling banks, investors, and regulators to benchmark performance, evaluate risk, and identify pathways toward more effective AI integration [7].

This research underscores the importance of data-driven frameworks, such as the Evident AI Index, for assessing technological evolution in the financial services industry. As AI redefines operational models and market dynamics, tools that provide comparative insight into adoption behavior are essential for guiding strategic foresight and digital capability development [8]. This study contributes to that endeavor by offering a rigorous analytical approach to understanding AI maturity and clustering within one of the most innovative and sensitive sectors of the global economy.

2. Methodology

This study systematically integrates PCA and k-means clustering to uncover patterns within the Evident AI Index dataset. This methodology provides a robust framework for dimensionality reduction, optimal cluster selection, and the interpretation of meaningful groupings.

2.1 Data Source

The primary dataset for this study was drawn from the 2024 Evident AI Index Rankings (https://evidentinsights.com/ai-index/), a structured benchmarking tool assessing the AI maturity of global banks. The Index evaluates institutions based on four key pillars: talent, innovation, leadership, and transparency. Each pillar comprises multiple indicators, yielding a multidimensional dataset that captures both qualitative and quantitative facets of AI adoption [4].

The Evident AI Index, created by Evident, is an intelligence platform that measures and monitors AI integration within the financial services industry. It offers an unparalleled standard for assessing AI adoption and development across the banking sector. It evaluates fifty of the largest banks in North America, Europe, and Asia using ninety indicators derived from millions of publicly available data points [4]. Table 1 presents descriptive statistics for the Evident AI index, providing a detailed overview of its underlying data.

Table 1. Descriptive Statistics of the Evident AI Index

N = 50	Mean	Std. Error	Std. Deviation
Talent	35.856	1.4944	10.5667
Innovation	23.430	1.9695	13.9262
Leadership	27.340	2.0818	14.7206
Transparency	30.878	2.7733	19.6099

2.2 Principal Component Analysis (PCA)

To ensure comparability across variables with different scales and units, all data were standardized using z-score normalization. This preprocessing step ensured that each feature contributed equally to the clustering process. Principal Component Analysis (PCA) was applied to reduce dimensionality and capture the most significant variance in the dataset. This process begins by computing the PCA components and analyzing the explained variance ratio to determine the optimal number of components to retain [9]. This step helped simplify the data structure and enhanced interpretability while mitigating multicollinearity [5].

2.3 K-Means Clustering

The reduced dataset from PCA was subjected to k-means clustering to identify distinct groups of banks based on their AI maturity profiles [10]. This involves iterating over a range of cluster counts and performing k-means for each specified number of clusters. This initial step generates multiple clustering outcomes, including cluster labels, centroids, and inertia values, which are further evaluated in subsequent steps [11]. The algorithm groups data points into clusters by minimizing the sum of squared errors (SSE) within clusters [12]. The optimal number of clusters was determined using the Elbow Method, Average Silhouette Score, Gap Statistic, and NbClust() Function. Each resulting cluster was then profiled based on its centroid characteristics to uncover adoption patterns. These complementary methods ensure a robust and consistent approach to determining the optimal cluster count.

3. Results and Discussion

3.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is employed to reduce the dimensionality of the dataset while preserving the variance that captures the primary factors influencing global AI competitiveness (Figure 1). This process involves calculating the principal components and examining the explained variance ratio to identify the optimal number of components to retain [9]. Table 2 provides a detailed understanding of how each original variable contributes to the principal components. In this table, each cell represents the loading of a variable on a principal component.

Table 2. Principal Component Analysis (PCA) Loadings

Index	PC1	PC2	PC3	PC4
Talent	-0.5301	-0.4799	-0.0019	0.6991
Innovation	-0.5335	-0.4340	0.1849	-0.7020
Leadership	-0.4553	0.6497	0.6001	0.1024
Transparency	-0.4765	0.3991	-0.7782	-0.0896

Eigenvalues quantify the variation captured by each principal component, serving as a measure of their importance. In this study, we analyzed the eigenvalues to determine the appropriate number of principal components to retain [9]. Table 3 presents the eigenvalues and the proportion of variance (i.e., information) explained by each principal component. The total variance of the dataset is equal to the sum of the eigenvalues, which is 4.

Table 3. Eigenvalues, Variance %, and Cumulative Variance %

Dim.	**Eigenvalue**	**Variance %**	**Cumulative Variance %**
Dim. 1	2.5668	64.1711	64.1711
Dim. 2	0.6674	16.6845	80.8556
Dim. 3	0.5098	12.7452	93.6008
Dim. 4	0.2560	6.3992	100.0000

The proportion of variation explained by each eigenvalue is presented in the second column. For instance, dividing an eigenvalue of 2.5668 by the total variance of 4 yields 0.6417, indicating that the first principal component explains approximately 64.17% of the variation. The cumulative percentage of variation explained is calculated by successively adding these proportions to obtain a running total. For example, adding 64.17% to 16.68% results in 80.85%. Thus, the first two eigenvalues account for approximately 80.85% of the total variation.

Vol. 6, No. 1, June 2025, pp. 29-39

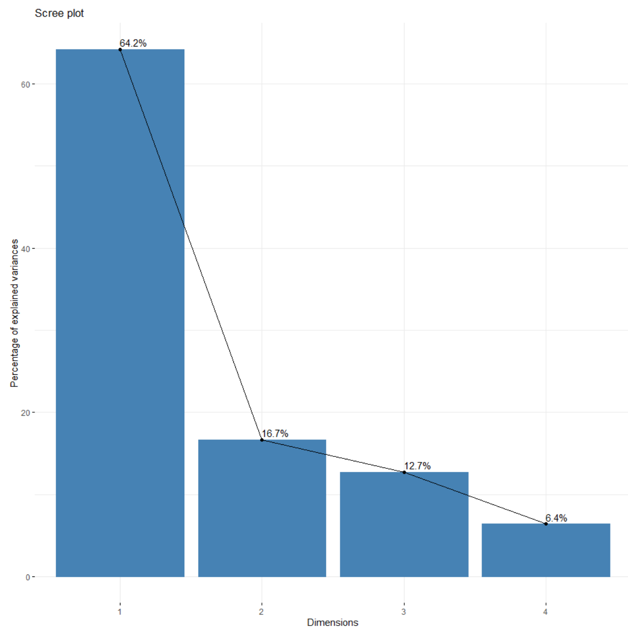

Figure 1. Screen Plot of Principal Components

The correlation between a variable and a principal component represents the variable's coordinate on that principal component. This type of visualization is known as a variable correlation plot (Figure 2), which illustrates the relationships among all variables. Additionally, a biplot (Figure 3) combines the variable correlations and observations, visually displaying the relationships between variables and individual data points. This dual representation enhances the interpretability of the principal components by showing how variables contribute to patterns within the observations.

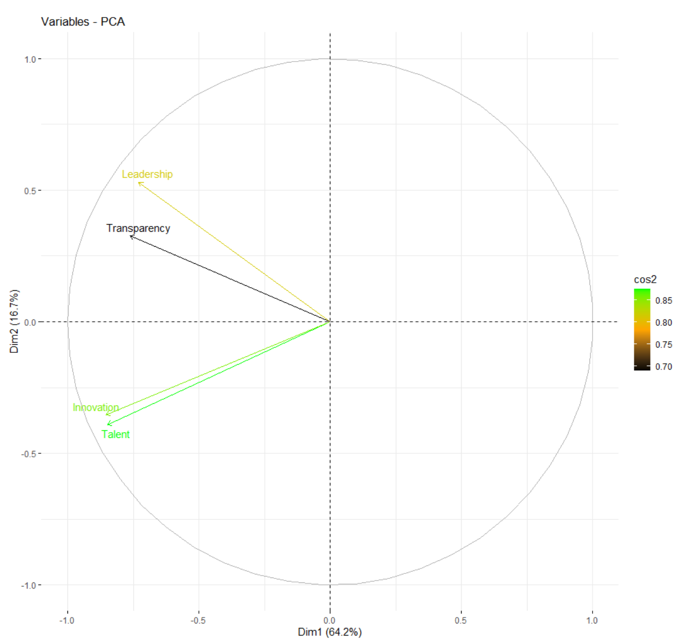

Figure 2. Variable Correlation Plot of Principal Components

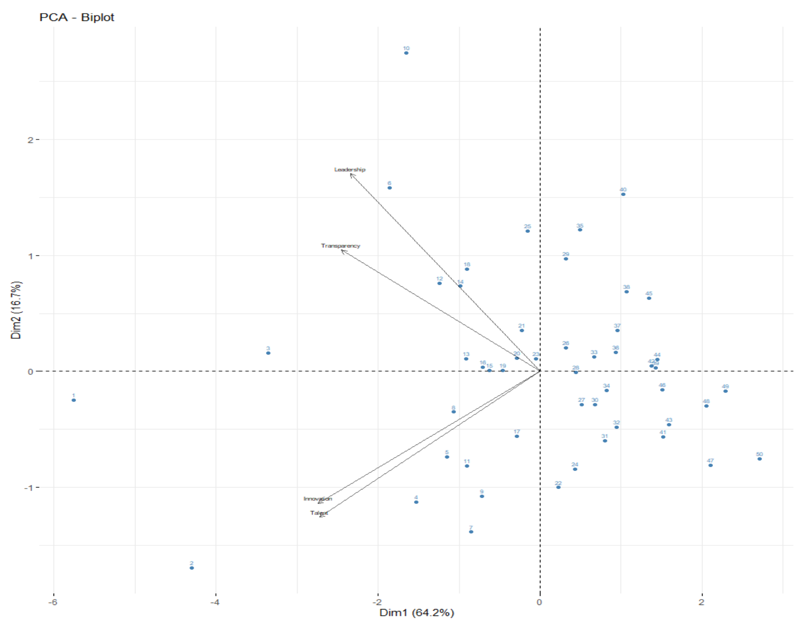

Figure 3. Biplot of Principal Components

3.2 K-Means Clustering

3.2.1 Elbow Method

The elbow method evaluates the total within-cluster sum of squares (WSS), also called total intra-cluster variation, as a function of the number of clusters. The WSS measures the compactness of the clustering, with smaller values indicating tighter clusters. The analysis of this study suggested an optimal solution with 3 clusters (Figure 4). However, it is essential to note that the elbow method can sometimes yield ambiguous results, as the bend in the plot is not always distinct. Alternatively, the average silhouette score method may be used. This method evaluates cluster quality based on cohesion and separation and can be applied to any clustering approach for additional validation.

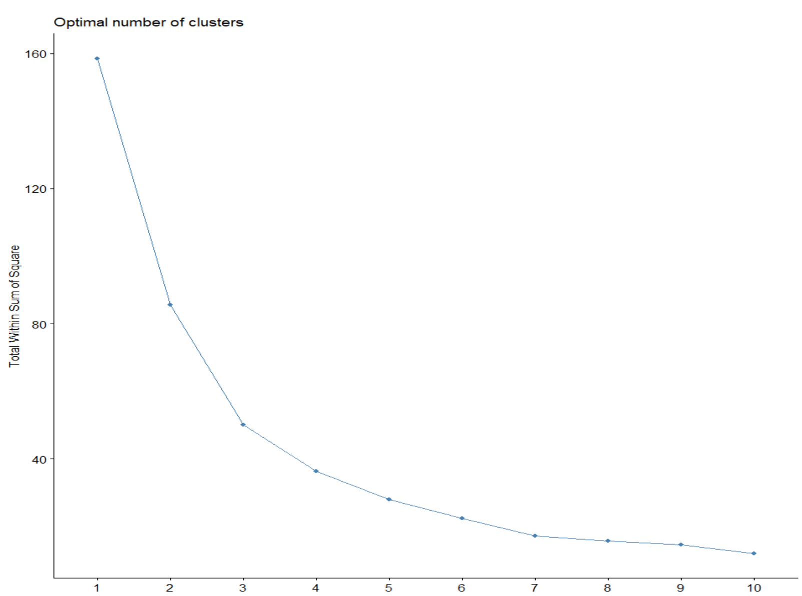

Figure 4. Optimal Number of Clusters Using the Elbow Method

3.2.1 Average Silhouette Score Method

The average silhouette score method evaluates clustering quality by computing the average silhouette score across different cluster values [13]. The optimal number of clusters is the one that maximizes the average silhouette score across a range of possible values. The silhouette score measures how well each point in a cluster is separated from points in neighboring clusters. A higher score indicates more defined and cohesive clusters. The silhouette plot visually represents this measure, providing insights into the clustering quality. The analysis of this study suggested an optimal solution with 3 clusters (Figure 5). This method complements the elbow method by offering a more quantitative and visually intuitive approach to determining the number of clusters.

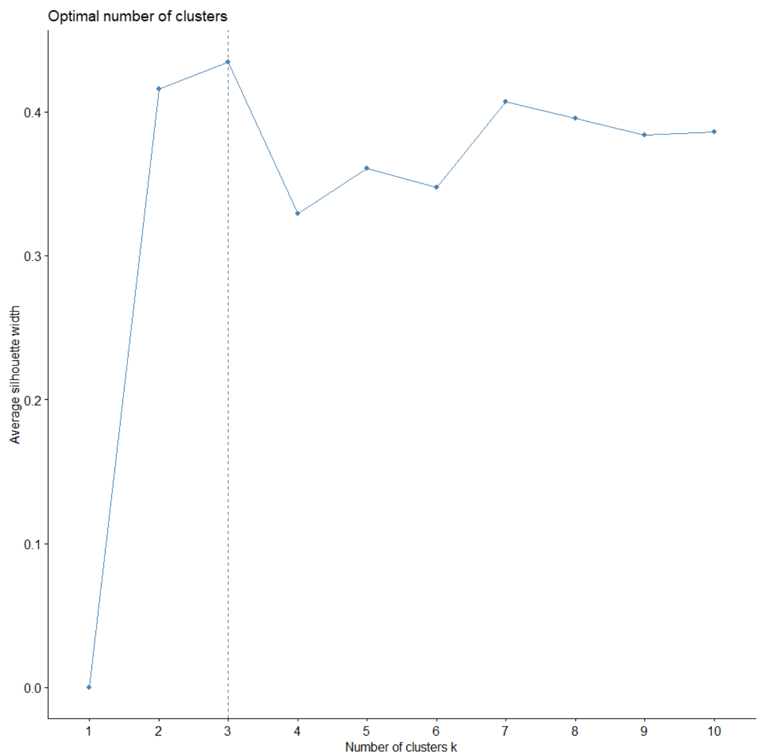

Figure 5. Optimal Number of Clusters Using the Average Silhouette Score Method

3.2.2 Gap Statistic Method

The gap statistic method evaluates the total within-cluster variation for various cluster counts and compares these values with their expected values under a null reference distribution of the data [14]. The optimal number of clusters is estimated as the value that maximizes the gap statistic (i.e., yields the most significant gap). This indicates that the clustering structure deviates significantly from a random, uniform distribution of points. While the elbow and average silhouette score methods provide valuable insights, they measure global clustering characteristics only and lack a formal statistical basis. The gap statistic method offers a more rigorous approach, formalizing the elbow/silhouette heuristics into a statistical procedure for estimating the optimal number of clusters. In this study, the analysis suggested an optimal solution of a single cluster (Figure 6), highlighting the statistical robustness of the gap statistic method.

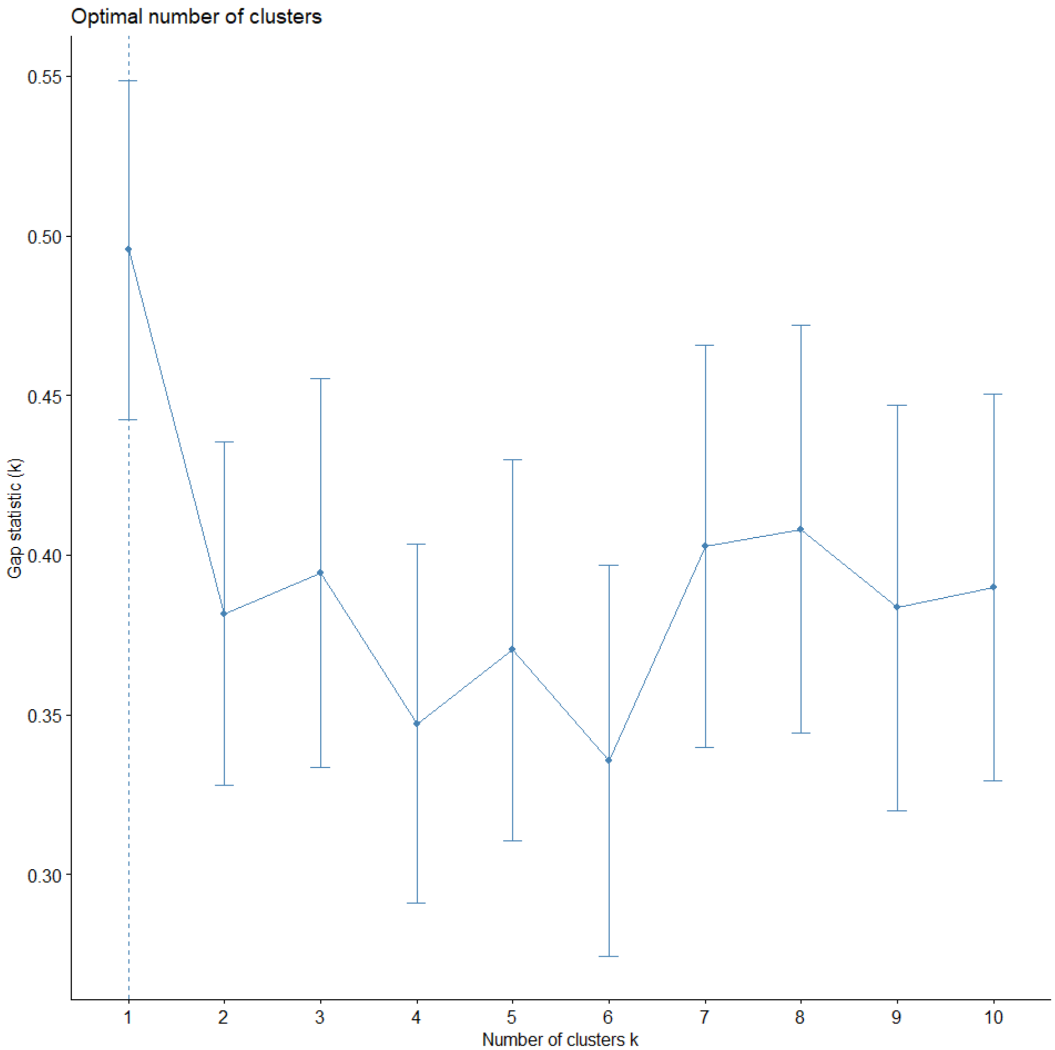

Figure 6. Optimal Number of Clusters Using the Gap Statistic Method

3.2.3 NbClust() Function

The NbClust() function, part of the NbClust package in R [15], is a powerful tool for determining the optimal number of clusters in clustering algorithms. It evaluates 30 indices to determine the optimal number of clusters and recommends the most appropriate clustering scheme based on the results. This function allows users to explore combinations of cluster count, distance measures, and clustering methods. In this study, based on the majority rule across the computed indices, the optimal number of clusters was determined to be three (Figure 7).

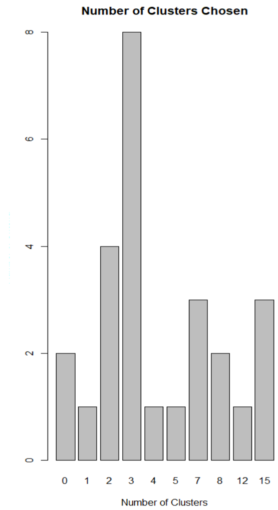

Figure 7. Optimal Number of Clusters Using NbClust() Function

3.3 Cluster Profiling

Following the clustering process, each cluster was meticulously analyzed in relation to the original variables: talent, innovation, leadership, and transparency. This analysis aimed to interpret the nature of AI adoption among the banks. Consequently, three distinct clusters emerged, as illustrated in Figures 8 and 9. These clusters were labeled based on their relative scores: "early AI adopters," "innovation leaders," and "conservative implementers."

Figure 8. Cluster Plot of the Three Identified Clusters

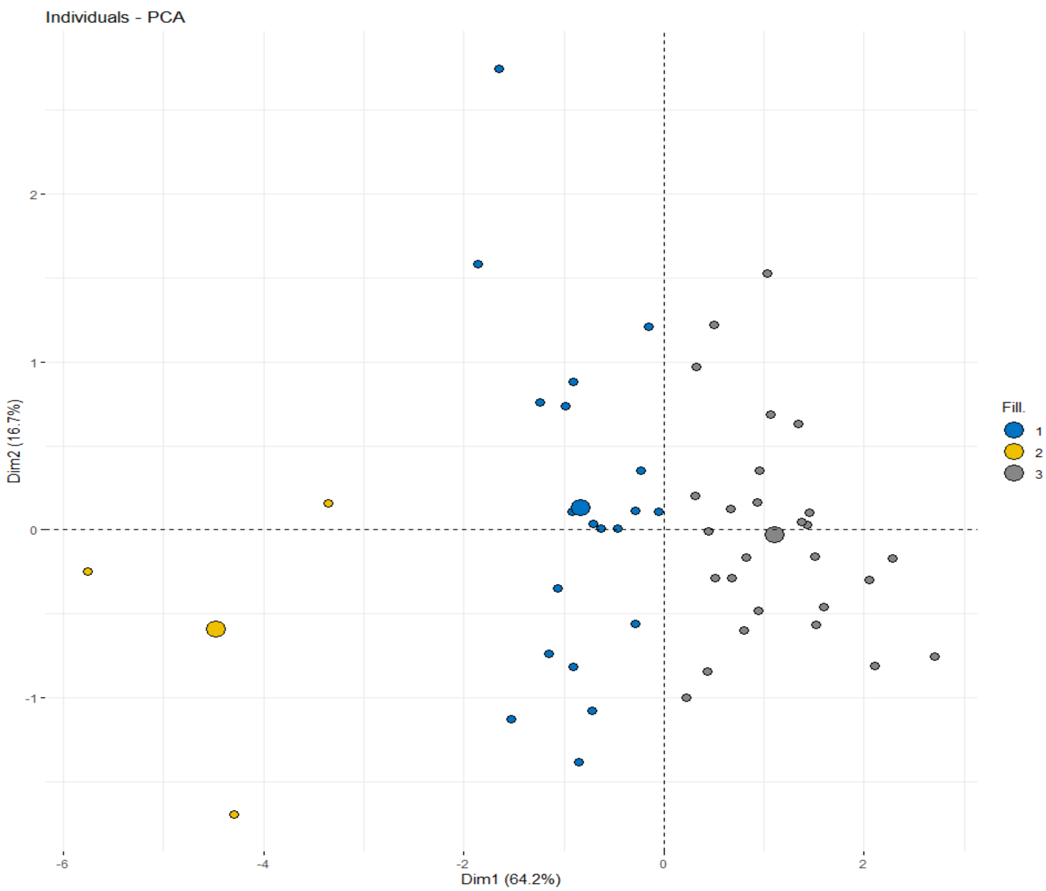

Figure 9. Individuals - PCA Plot of the Three Identified Clusters

The ANOVA results for the Evident AI Index reveal significant differences between clusters across all four indices: Talent, Innovation, Leadership, and Transparency (Table 6). For the Talent Index, the F-value of 35.178 and a significance level of < 0.001 indicate substantial variability between groups. Similarly, the Innovation Index shows an F-value of 69.009 with a significant level of < 0.001, highlighting notable differences among clusters. The Leadership Index also demonstrates significant differences, with an F-value of 15.372 and a significance level of < 0.001. Lastly, the Transparency Index presents an F-value of 21.870 and a significance level of < 0.001, confirming significant disparities between groups. These results underscore each cluster's distinct characteristics and performance levels, suggesting that the clusters are meaningfully differentiated in terms of Talent, Innovation, Leadership, and Transparency.

Table 6. ANOVA Table

Index		**Sum of Squares**	***df***	**Mean Square**	***F***	**Sig.**
Talent	Between Groups	3279.972	2	1639.986	35.178	< 0.001
	Within Groups	2191.151	47	46.620		
	Total	5471.123	49			
Innovation	Between Groups	7088.946	2	3544.473	69.009	< 0.001
	Within Groups	2414.019	47	51.362		
	Total	9502.965	49			
Leadership	Between Groups	4198.902	2	2099.451	15.372	< 0.001
	Within Groups	6419.158	47	136.578		
	Total	10618.060	49			
Transparency	Between Groups	9082.887	2	4541.443	21.870	< 0.001
	Within Groups	9760.059	47	207.661		
	Total	18842.946	49			

4. Conclusion

This study investigated the adoption patterns of artificial intelligence in the global banking sector using a PCA-based k-means clustering methodology, drawing from the Evident AI Index Rankings. Through dimensionality reduction and unsupervised learning, the analysis identified three primary segments: early adopters, innovation leaders, and conservative implementers. These clusters revealed significant variation in how banks approach AI across talent acquisition, innovation, leadership visibility, and AI transparency.

The emergence of early adopters, innovation leaders, and conservative implementers aligns with findings in digital transformation literature, which suggest that strategic priorities and institutional readiness significantly affect the pace and scope of AI integration [16], [8]. The innovation leaders group demonstrates that sustained investments in AI talent and R&D correlate with broader leadership in digital transformation. These banks will likely capture early-mover advantages, including efficiency gains, customer retention, and regulatory readiness. Conversely, conservative implementers may face competitive pressures and risk-management challenges, particularly in an era increasingly driven by data and automation.

The clustering results have several policy implications for regulators and industry stakeholders. First, the uneven distribution of AI maturity underscores the need for differentiated regulatory engagement. Innovation leaders may benefit from regulatory sandboxes that enable experimentation, whereas conservative implementers may require structured guidance to adopt AI safely within a compliance framework. Second, the study highlights the role of transparency as a signaling mechanism. Banks that publicly disclose AI strategies, ethical commitments, and performance metrics tend to rank higher in maturity. Regulators and oversight bodies could encourage standardized AI reporting practices, fostering greater accountability and trust in AI systems.

Finally, the importance of AI talent concentration underscores the need for cross-sector collaboration among governments, academia, and industry to build inclusive talent pipelines. The talent gap could exacerbate global inequalities in AI adoption and financial innovation without such initiatives. By combining the analytical strength of PCA with the interpretive clarity of clustering, this research offers a structured perspective on the relative AI maturity of institutions. The segmentation uncovers current competitive positioning and provides a foundation for tracking technological trajectories over time. Moreover, using PCA and k-means proved effective in uncovering nuanced adoption patterns that would be difficult to detect using traditional categorical assessments. Integrating benchmarking tools like the Evident AI Index into clustering analysis provides a replicable framework for tracking technological transformation over time.

The study contributes to the growing body of literature seeking to demystify AI deployment in financial services, offering actionable insights for banks, regulators, and policymakers. From a policy and strategic planning perspective, this segmentation can guide stakeholders in crafting tailored interventions, supporting lagging banks through targeted talent programs or fostering collaborative ecosystems that incentivize AI adoption across the financial sector.

References

[1] J. Bughin, J. Seong, J. Manyika, M. Chui, and R. Joshi, *Notes from the AI Frontier: Modeling the Impact of AI on the World Economy*. New York, NY, USA: McKinsey Global Institute, 2018.

[2] E. Brynjolfsson and A. McAfee, *Machine, Platform, Crowd: Harnessing Our Digital Future*. New York, NY, USA: W.W. Norton & Company, 2017.

[3] D. W. Arner, J. Barberis, and R. P. Buckley, "FinTech, RegTech, and the reconceptualization of financial regulation," *Northwestern J. Int. Law Business*, vol. 37, no. 3, pp. 371–414, 2017. [Online]. Available: https://scholarlycommons.law.northwestern.edu/njilb/vol37/iss3/2.

[4] Evident, *The Evident AI Index Banks: The Global Standard Benchmark of AI Maturity*. London, U.K.: Evident Insights, 2024.

[5] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philosophical Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 374, no. 2065, p. 20150202, 2016, doi: 10.1098/rsta.2015.0202.

[6] S. Makridakis, "The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms," *Futures*, vol. 90, pp. 46–60, Jun. 2017, doi: 10.1016/j.futures.2017.03.006.

[7] L. Zavolokina, M. Dolata, and G. Schwabe, "The FinTech phenomenon: Antecedents of financial innovation perceived by the popular press," *Financial Innov.*, vol. 2, no. 1, p. 16, Dec. 2016, doi: 10.1186/s40854-016-0036-7.

[8] A. Belelieu, D. Propson, and D. Parker, "Artificial intelligence in financial services," World Econ. Forum, White Paper, 2025.

[9] A. Kassambara, *Practical Guide to Principal Component Methods in R: PCA, (M)CA, FAMD, MFA, HCPC, Factoextra (Multivariate Analysis)*, 1st ed. U.S.A.: STHDA, 2017.

[10] A. Kassambara, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning (Multivariate Analysis)*, 1st ed. U.S.A.: STHDA, 2017.

[11] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/j.patrec.2009.09.011.

[12] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, Berkeley, CA, USA, 1967, vol. 1, pp. 281–297.

[13] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ, USA: John Wiley & Sons, 2005.

[14] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *J. Roy. Statist. Soc., Ser. B (Statist. Methodol.)*, vol. 63, no. 2, pp. 411–423, 2001, doi: 10.1111/1467-9868.00293.

[15] M. Charrad, N. Ghazzali, V. Boiteau, and A. Niknafs, "NbClust: An R package for determining the relevant number of clusters in a data set," *J. Statist. Softw.*, vol. 61, no. 6, pp. 1–36, 2014, doi: 10.18637/jss.v061.i06.

[16] A. Fuster, P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther, "Predictably unequal? The effects of machine learning on credit markets," *J. Finance*, vol. 77, no. 1, pp. 5–47, Feb. 2022, doi: 10.1111/jofi.13090.