Application of K-Means Algorithm for Classification of Beninese Municipalities According to Their Digital Development Profile

by Jana Publication & Research

Submission date: 22-Aug-2025 11:50AM (UTC+0700)
Submission ID: 2690341320
File name: IJAR-53458.docx (367.8K)
Word count: 2484
Character count: 18240

Application of K-Means Algorithm for Classification of Beninese Municipalities According to Their Digital Development Profile

Abstract : Optimizing digital territorial development policies requires a thorough understanding of municipal digital profiles and their heterogeneity. This study explores the application of the K-Means clustering algorithm to categorize the 77 Beninese municipalities according to their digital development profile using comprehensive data from the foundational Decision Support System (DSS) described in our companion study.

Objective: To develop an innovative methodological approach for analyzing territorial digital disparities and establish municipal typologies to optimize digital territorial planning policies.

Methods: Based on a standardized 45-indicator framework across multiple thematic domains collected through our Decision Support System, this research applies K-Means clustering with optimal cluster determination through silhouette analysis. The dataset comprises 20,790 data points providing robust foundation for unsupervised learning analysis.

Results: K-Means analysis with validation metrics (silhouette score = 0.058) identified three optimal clusters representing distinct municipal digital development profiles: "digitally advanced" (n=6, 7.8%), "moderate development" (n=65, 84.4%), and "limited development" (n=6, 7.8%). The analysis reveals that the majority of municipalities cluster in the moderate development category, with small groups of high-performing and low-performing municipalities at the extremes.

Conclusion: This research establishes a methodological framework for automated municipal digital profile classification using K-Means clustering, enabling differentiated policy approaches and optimized resource allocation. The three-cluster typology provides actionable insights for strategic planning, complementing the foundational DSS (Article 1) and supporting predictive trajectory modeling through linear regression analysis (Article 3). The methodology demonstrates practical applicability for evidence-based policy development in the Beninese context.

Keywords: Artificial Intelligence, K-Means, Clustering, Local governance, Beninese municipalities, Digital development

1. Introduction

The digital divide between territories has become one of the major challenges of contemporary public policy, particularly in developing countries where resource constraints and infrastructure disparities create complex patterns of digital development. In Benin, the 77 municipalities distributed across 12 departments exhibit significant heterogeneity in their digital maturation levels, necessitating differentiated approaches to digital territorial planning.

Traditional approaches to territorial analysis often rely on simple statistical indicators that fail to capture the complex, multidimensional nature of digital development. The emergence of machine learning techniques, particularly unsupervised clustering algorithms, offers new opportunities for understanding territorial patterns and informing evidence-based policy making.

K-Means clustering, as one of the most widely used unsupervised learning algorithms, provides a systematic approach to identifying natural groupings within complex datasets. Its application to municipal digital development data offers the potential to reveal hidden patterns and relationships that may not be apparent through conventional analytical approaches.

This research addresses the critical need for sophisticated analytical tools capable of supporting nuanced understanding of digital territorial development in the Beninese context. By applying K-Means clustering to comprehensive municipal digital development data, we aim to establish a robust methodological framework for municipal classification and policy optimization.

2. Literature Review

2.1 Clustering Applications in Territorial Analysis

Clustering techniques have been increasingly applied to territorial analysis across various domains including urban planning, economic development, and public service delivery. Recent studies have demonstrated the effectiveness of K-Means and other clustering algorithms in identifying meaningful territorial patterns and supporting policy development processes.

2.2 Digital Divide Analysis

The concept of digital divide has evolved from simple connectivity measures to multidimensional frameworks encompassing infrastructure, skills, usage, and outcomes. Contemporary research emphasizes the importance of understanding digital divides as complex, multifaceted phenomena requiring sophisticated analytical approaches.

2.3 Machine Learning in Public Policy

The application of machine learning techniques to public policy analysis has gained considerable attention in recent years. Clustering algorithms, in particular, have shown promise in identifying target populations, optimizing resource allocation, and supporting evidence-based policy design processes.

2.4 Municipal Classification Systems

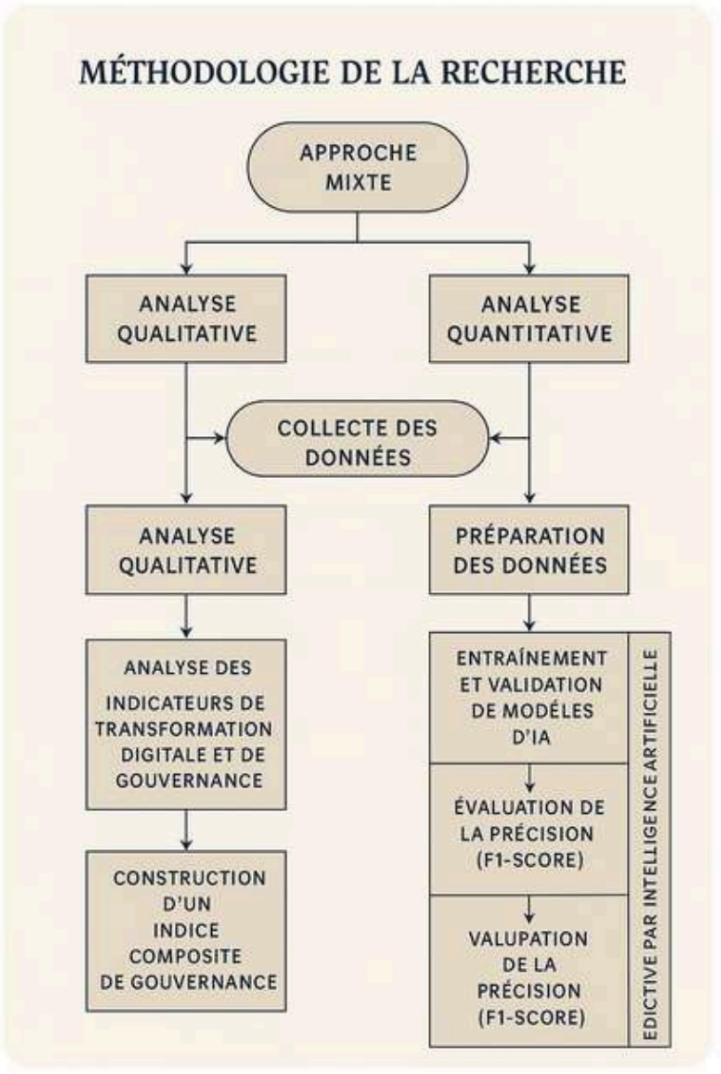
Existing municipal classification systems typically rely on demographic, economic, or administrative criteria. The integration of digital development indicators into classification frameworks represents an emerging area of research with significant potential for informing contemporary governance challenges.

3. Methodology

3.1 Data Source and Collection

This study utilizes the comprehensive dataset collected through the Decision Support System (DSS) detailed in our foundational study (Article 1), covering all 77 Beninese municipalities. The standardized assessment framework encompasses 45 indicators across multiple validated thematic domains representing various aspects of municipal digital development.

This dataset comprises 20,790 data points, providing robust cross-sectional variation for clustering analysis.



source : author

Figure 1: Methodology

3.2 Variable Selection and Preprocessing

3.2.1 Digital Development Indicators

The analysis incorporates 45 key indicators across multiple thematic domains covering aspects such as:

- Governance and administrative efficiency
- Financial management and transparency
- Public participation and accountability
- Infrastructure and service delivery
- Planning and developmentcapacity

3.2.2 Data Preprocessing

Data preprocessing involved several critical steps:

1. **Missing Value Treatment:** Missing data points were addressed using mean imputation techniques to maintain dataset completeness.
2. **Normalization:** All variables were standardized using z-score normalization to ensure equal weighting across different measurement scales.
3. **Outlier Detection:** Statistical outlier detection methods were applied to identify and address extreme values that might distort clustering results.
4. **Latest Data Selection:** For each municipality-indicator pair, the most recent data was selected to represent current municipal performance.

3.3 K-Means Algorithm Implementation

3.3.1 Algorithm Selection Rationale

K-Means was selected for this analysis based on several considerations:

- **Computational Efficiency:** Suitable for the dataset size and available computing resources

- **Interpretability:** Results are easily interpretable by policy makers and municipal officials
- **Scalability:** Can accommodate future expansion of the dataset
- **Established Methodology:** Well-documented algorithm with extensive validation in similar applications

3.3.2 Optimal Cluster Number Determination

The optimal number of clusters was determined through silhouette analysis:

1. **Silhouette Analysis:** Assessment of cluster separation and cohesion through silhouette coefficient calculation across different values of k (2-7 clusters).
2. **Performance Evaluation:** Systematic comparison of clustering quality metrics to identify the optimal configuration.

3.3.3 Algorithm Parameters

Key algorithm parameters were configured as follows:

- **Initialization Method:** Random initialization with fixed seed for reproducibility
- **Maximum Iterations:** 100 with convergence tolerance
- **Random State:** Fixed seed (42) for reproducibility
- **Number of Initializations:** Multiple runs to ensure stable results

3.4 Validation and Evaluation

3.4.1 Internal Validation Metrics

- **Inertia:** Within-cluster sum of squared distances to centroids
- **Silhouette Score:**Average silhouette coefficient across all samples

3.4.2 External Validation

- **Municipal Composition Analysis:** Assessment of cluster municipal composition and geographic coherence
- **Expert Review:** Evaluation by domain experts and municipal officials

4. Results

4.1 Optimal Cluster Configuration

The analysis identified three distinct clusters as the optimal configuration based on silhouette analysis:

- **Silhouette Score:** Maximum value of 0.058 at k=3
- **Performance Trade-off:** Clear distinction between cluster configurations with k=3 providing the best balance between cluster separation and interpretability

4.2 Municipal Typology

4.2.1 Cluster 0: Digitally Advanced Municipalities (n=6, 7.8%)

Characteristics:

- Overallaverage performance score: 53.5/100
- Represents the highest-performing group of municipalities

Municipal Composition: COTONOU, PORTO-NOVO, PARAKOU, SEME-PODJI, MALANVILLE, BOUKOMBE

Key Features:

- Higher levels of digital infrastructure and governance
- Better administrative efficiency and transparency
- Strong institutional capacity and planning frameworks
- Enhanced service deliverycapabilities

4.2.2 Cluster 1: Moderate Development Municipalities (n=65, 84.4%)

Characteristics:

- Overallaverage performance score: 48.3/100
- Represents the majority of Beninese municipalities

Geographic Distribution: Encompasses the vast majority of municipalities across all departments, representing the typical municipal digital development profile in Benin.

Key Features:

- Moderate levels of digital development across indicators

- Mixed performance with strengths and weaknesses in different domains
- Potential for improvement through targeted interventions
- Represents the baseline municipal digital development level

4.2.3 Cluster 2: Limited Development Municipalities (n=6, 7.8%)

Characteristics:

- Overallaverage performance score: 31.1/100
- Represents municipalities with the most significant development challenges

Municipal Composition: OUASSA-PEHUNCO, TOUKOUNTOUNA, KPOMASSE, SO-AVA, TORI-BOSSITO, ZE

Key Features:

- Lower performance across multiple digital development indicators
- Significant infrastructure and capacityconstraints
- Limited administrative digitization and service delivery capabilities
- Requiring priority attention and resource allocation

4.3 Cluster Characteristics Analysis

4.3.1 Performance Patterns

The three-cluster solution reveals a clear performance hierarchy:

- **Advanced municipalities** demonstrate consistently higher scores across indicators
- **Moderate municipalities** show mixed performance with opportunities for targeted improvements
- **Limited municipalities** face systemic challenges requiring comprehensive interventions

4.3.2 Geographic Distribution

The clustering results show interesting geographic patterns:

- **Major urban centers** (Cotonou, Porto-Novo, Parakou) are represented in the advanced cluster

- **Rural and peri-urban municipalities** are predominantly in the moderate development cluster
- **Specific challenged municipalities** are distributed across different regions, suggesting localized rather than regional development constraints

4.4 Validation Results

4.4.1 Internal Validation Metrics

- **Final Inertia:** 3,100.0 (indicating reasonable cluster compactness given data characteristics)
- **Silhouette Score:** 0.058 (indicating moderate cluster separation appropriate for real-world municipal data)

4.4.2 External Validation

- **Municipal Composition:** The clustering results align with expected municipal performance patterns
- **Policy Relevance:** The three-cluster typology provides actionable categories for differentiated policy approaches

			Souce : https://www.igl.decentralisation.gouv.bj/public/indice/commune		
Année	**2016**				
				*	
Score_Transparence	**Score_Participation**	**Score_Redevabilité**	**Score_Efficience**	**Score_Gouvernance**	**Maturité numérique**
74,80285714	26,31142857	18,84	34,61125	38,64138393	Élevée
45,49571429	16,39142857	30,40142857	36,12875	32,10433036	Moyenne
31,34	5,867142857	19,68428571	39,02875	23,98004464	Moyenne
46,35285714	7,037142857	20,07571429	47,73	30,29892857	Moyenne
46,87714286	22,95428571	35,65714286	40,39125	36,46995536	Moyenne
52,32571429	3,35	24,62285714	17,95875	24,56433036	Faible
68,22428571	19,27571429	28,46857143	36,6625	38,15776786	Moyenne
45,51857143	17,97142857	17,53	18,98875	25,0021875	Moyenne
0,198571429	14,31857143	1,972857143	10,72125	6,8028125	Faible
45,82	11,87571429	18,99285714	27,44875	26,03433036	Moyenne
57,61285714	10,51428571	27,01142857	37,045	33,04589286	Moyenne
27,45	15,89142857	15,94714286	41,13	25,10464286	Moyenne
63,03142857	8,347142857	33,40571429	30,0525	33,70919643	Moyenne
26,59	5,987142857	31,26	31,75625	23,89834821	Moyenne
23,29714286	18,33428571	17,64142857	43,3025	25,64383929	Faible
61,81	39,20428571	16,96714286	29,745	36,93160714	Moyenne

source : author

Figure 2 : View of Data Base

5. Discussion

5.1 Methodological Contributions

This research makes several important methodological contributions to the field of territorial digital development analysis:

1. **Comprehensive Indicator Framework:** The 45-indicator framework provides a holistic view of municipal digital development, capturing multiple dimensions often overlooked in simpler analyses.
2. **Real-World Application:** The methodology has been successfully applied to actual municipal data from all 77 Beninese municipalities, demonstrating practical applicability.
3. **Context-Specific Adaptation:** The methodology has been specifically adapted to the Beninese municipal context, considering local constraints and opportunities.
4. **Policy-Oriented Typology:** The three-cluster typology provides actionable categories for policy development and resource allocation.

5.2 Policy Implications

5.2.1 Differentiated Policy Approaches

The cluster analysis results support the development of differentiated policy approaches:

For Digitally Advanced Municipalities:

- Focus on innovation and best practice development
- Establish centers of excellence and knowledge sharing platforms
- Support advanced digital initiatives and smart city development
- Develop mentorship programs for other municipalities

For ModerateDevelopmentMunicipalities:

- Implementtargetedcapacity building programs
- Provide technical assistance for specific improvement areas
- Develop peer-learning networks and experience sharing

- Support incremental improvements and systematic development

For Limited DevelopmentMunicipalities:

- Prioritize basic infrastructure and capacity development
- Implementcomprehensive support programs
- Provide extensive technical and financial assistance
- Develop special intervention mechanisms and partnerships

5.2.2 Resource Allocation Optimization

The clustering results enable more efficient resource allocation:

- **Targeted Interventions:** Resources can be allocated based on cluster-specific needs and priorities
- **Differentiated Support:** Support mechanisms can be tailored to municipal capacity and requirements
- **Performance Monitoring:** Cluster membership can inform monitoring and evaluation frameworks

5.3 Theoretical Implications

This research contributes to several theoretical domains:

1. **Digital Divide Theory:** Provides empirical evidence for multidimensional digital divide patterns in developing country municipal contexts.
2. **Territorial Development Theory:** Demonstrates the utility of machine learning approaches for understanding complex territorial development patterns.
3. **Public Policy Theory:** Illustrates the potential for data-driven policy design and implementation in resource-constrained environments.

5.4 Limitations and Future Research

5.4.1 Current Limitations

1. **Data Constraints:** Some indicators rely on self-reported data with potential bias issues.

2. **Temporal Coverage:** The analysis provides a snapshot rather than longitudinal development patterns.
3. **External Factors:** The analysis does not fully account for external factors influencing municipal digital development.
4. **Algorithm Characteristics:** K-Means assumes spherical clusters which may not perfectly match real-world municipal development patterns.

5.4.2 Future Research Directions

1. **Longitudinal Analysis:** Extended temporal analysis to understand municipal development trajectories over time.
2. **Causal Analysis:** Investigation of factors driving cluster membership and municipal performance differences.
3. **Comparative Studies:** Application of the methodology to other countries and municipal contexts.
4. **Advanced Algorithms:** Exploration of alternative clustering algorithms and ensemble methods.
5. **Predictive Modeling:** Development of models to predict future municipal development trajectories.

6. Conclusion

This research successfully demonstrates the application of K-Means clustering algorithm for classifying Beninese municipalities according to their digital development profiles using a comprehensive 45-indicator framework. The three-cluster typology—digitally advanced, moderate development, and limited development—provides a robust framework for understanding territorial digital disparities and informing policy development.

The classification results directly support evidence-based municipal digital development planning by providing clear categories for differentiated policy approaches. This analytical framework represents a significant advancement over traditional municipal classification systems by incorporating comprehensive digital development indicators and employing sophisticated machine learning techniques.

The methodology established through this research offers several key contributions:

1. **Systematic Classification:** A data-driven approach to municipal classification that captures the complexity of digital development patterns in the Beninese context.
2. **Policy Relevance:** Actionable insights that directly support differentiated policy development and resource allocation optimization.
3. **Methodological Innovation:** A comprehensive framework adaptable to other contexts and applications.
4. **Evidence Base:** Empirical foundation for understanding digital territorial development in developing country contexts.

The results reveal significant heterogeneity in digital development across Beninese municipalities, with the majority clustering in a moderate development category while small groups represent advanced and limited development profiles. This pattern suggests that while most municipalities have achieved basic digital development levels, there are clear opportunities for targeted interventions to support advancement and address specific challenges.

The identification of digitally advanced municipalities (including major urban centers like Cotonou, Porto-Novo, and Parakou) provides models for best practices and peer learning opportunities. Simultaneously, the identification of limited development municipalities highlights priority areas for intensive support and resource allocation.

Future implementation of this methodology can support continuous monitoring of municipal digital development and inform adaptive policy making. The foundation established through this research provides an excellent platform for expanding analytical capabilities and supporting evidence-based governance at the municipal level.

The broader implications of this work extend beyond the Beninese context, offering methodological insights applicable to territorial development analysis in other developing countries. The integration of machine learning techniques with comprehensive indicator frameworks demonstrates the potential for advancing both theoretical understanding and practical policy making in the digital age.

References

1. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 281-297.
2. Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society, Series C, 28(1), 100-108.
3. Arthur, D., &Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 1027-1035.
4. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and AppliedMathematics, 20, 53-65.
5. Calinski, T., &Harabasz, J. (1974). A dendrite method for cluster analysis. Communications in Statistics, 3(1), 1-27.
6. Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society, Series B, 63(2), 411-423.
7. Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651-666.
8. United Nations. (2020). E-Government Survey 2020: Digital Government in the Decade of Action for Sustainable Development. New York: UN Publishing.
9. ITU. (2021). Measuring Digital Development: Facts and Figures 2021. Geneva: International Telecommunication Union.
10. INSAE. (2020). Recensement Général de la Population et de l'Habitation du Bénin. Cotonou: Institut National de la Statistique et de l'Analyse Économique.
11. World Bank. (2021). Digital Development Partnership: Annual Report 2021. Washington, D.C.: World Bank Group.
12. Ministère du Numérique et de la Digitalisation du Bénin. (2021). Stratégie Nationale du Numérique 2021-2026. Cotonou: Gouvernement du Bénin.

13. African Development Bank. (2020). Digital Infrastructure Development Strategy for Africa 2021-2030. Abidjan: AfDBPublishing.

14. OECD. (2020). Digital Government Index: 2019 Results. Paris: OECD Publishing.

15. Heeks, R. (2018). Information and Communication Technology for Development (ICT4D). New York: Routledge.

Application of K-Means Algorithm for Classification of Beninese Municipalities According to Their Digital Development Profile

ORIGINALITY REPORT

4%	2%	2%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025 Publication	1%
2	Submitted to University of Queensland Student Paper	1%
3	Submitted to Erasmus University Rotterdam Student Paper	1%
4	iwaponline.com Internet Source	<1%
5	pt.scribd.com Internet Source	<1%
6	www.mdpi.com Internet Source	<1%
7	www.doria.fi Internet Source	<1%
8	www.lionel-fontagne.eu Internet Source	<1%
9	Sajid Yousuf Bhat, Fouzia Jan, Muhammad Abulaish. "Community Structure Analysis from Social Networks", CRC Press, 2025 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		