Conciencia Digital

ISSN: 2600-5859
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original

www.concienciadigital.org

Bayesian nonparametric models in market segmentation: a systematic review and research agenda

Modelos bayesianos no paramétricos en la segmentación de mercados: una revisión sistemática y agenda de investigación

[1] Tannia Carolina Calle Monroy https://orcid.org/0009-0009-1180-4825
Universidad de Cuenca (UCUENCA), Cuenca, Ecuador.
tannia.monroy@ucuenca.edu.ec

[2] Saúl Fernando Pesántez Vicuña https://orcid.org/0000-0001-6861-3151
Universidad de Cuenca (UCUENCA), Cuenca, Ecuador.
fpesantez@gmail.com

[3] César Mesías Izquierdo Galarza https://orcid.org/0009-0005-6744-0339
Universidad de Cuenca (UCUENCA), Cuenca, Ecuador.
cesarizquierdog@gmail.com

[4] Denise Liliana Pazmiño Garzón https://orcid.org/0000-0002-9733-7114
Universidad de Cuenca (UCUENCA), Cuenca, Ecuador.
denise.pazminog@ucuenca.edu.ec

Artículo de Investigación Científica y Tecnológica
Enviado: 15/04/2026
Revisado: 10/05/2026
Aceptado: 26/06/2026
Publicado: 17/07/2026
DOI: **https://doi.org/10.33262/concienciadigital.v9i3.3704**

Cítese: Calle Monroy, T. C., Pesántez Vicuña, S. F., Izquierdo Galarza, C. M., & Pazmiño Garzón, D. L. (2026). Modelos bayesianos no paramétricos en la segmentación de mercados: una revisión sistemática y agenda de investigación. ConcienciaDigital, 9(3), 28-59. https://doi.org/10.33262/concienciadigital.v9i3.3704

CONCIENCIA DIGITAL, *es* una revista multidisciplinar, **trimestral,** que se publicará en soporte electrónico tiene como **misión** contribuir a la formación de profesionales competentes con visión humanística y crítica que sean capaces de exponer sus resultados investigativos y científicos en la misma medida que se promueva mediante su intervención cambios positivos en la sociedad. https://concienciadigital.org
La revista es editada por la Editorial Ciencia Digital (Editorial de prestigio registrada en la Cámara Ecuatoriana de Libro con No de Afiliación 663) www.celibro.org.ec

Palabras claves: Bayesian nonparametrics, market segmentation, clustering, Dirichlet process, systematic review.

Resumen

Introduction: Bayesian nonparametric (BNP) models have emerged as flexible approaches for clustering and segmentation by allowing data-driven identification of latent structures. Despite their methodological advances, their application in marketing remains limited. **Objective:** To systematically examine the literature on Bayesian nonparametric (BNP) models for clustering and segmentation, with a focus on their applicability in marketing contexts. **Methodology:** A PRISMA-based protocol was employed to identify and analyze 43 Scopus-indexed peer-reviewed articles. Bibliometric and thematic analyses were conducted to characterize research trends and methodological developments. **Results:** The results indicate a marked growth in scientific output, concentrated in statistical and methodological journals, with a strong emphasis on Dirichlet processes, mixture models, and clustering techniques. A taxonomy is proposed that structures the literature into six categories, ranging from foundational BNP models to advanced hierarchical and data-adaptive extensions. **Conclusions:** The findings reveal a persistent gap between methodological advances and their application in marketing, where traditional clustering approaches prevail. BNP models remain underutilized despite their capacity to capture latent heterogeneity and emergent segmentation structures. **General study area:** Marketing. **Specific study area:** Market segmentation, Bayesian nonparametric models, clustering. **Article type:** Systematic literature review.

Keywords: Bayesiano no paramétrico, segmentación de mercados, clustering, proceso de Dirichlet, revisión sistemática.

Abstract

Introducción: Los modelos bayesianos no paramétricos (BNP) han emergido como enfoques flexibles para el clustering y la segmentación al permitir la identificación de estructuras latentes guiadas por los datos. A pesar de sus avances metodológicos, su aplicación en contextos de marketing sigue siendo limitada. **Objetivo:** Examinar sistemáticamente la literatura sobre modelos bayesianos no paramétricos (BNP) aplicados al clustering y la segmentación, con énfasis en su aplicabilidad en contextos de marketing. **Metodología:** Se empleó un protocolo PRISMA para identificar y analizar 43 artículos revisados por pares indexados en Scopus. Se realizaron análisis bibliométricos y temáticos para caracterizar las tendencias de investigación y los desarrollos metodológicos. **Resultados:** Los resultados evidencian un

Conciencia Digital
ISSN: 2600-5859
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original
www.concienciadigital.org

crecimiento sostenido de la producción científica, concentrada principalmente en revistas estadísticas y metodológicas, con énfasis en procesos de Dirichlet, modelos de mezcla y técnicas de clustering. Se propone una taxonomía que estructura la literatura en seis categorías, desde modelos fundamentales hasta extensiones jerárquicas y adaptativas. **Conclusiones:** Los hallazgos muestran una brecha persistente entre los avances metodológicos y su aplicación en marketing, donde predominan enfoques tradicionales, pese a la capacidad de los modelos BNP para capturar heterogeneidad latente. **Área de estudio general:** Marketing. **Área de estudio específica:** Segmentación de mercados, modelos bayesianos no paramétricos, clustering. **Tipo de artículo**: Revisión sistemática de la literatura.

1. Introduction

Bayesian nonparametric (BNP) models have become a robust framework for clustering and modeling latent heterogeneity, primarily due to their ability to infer the number of clusters from the data and to accommodate complex, high-dimensional structures. Approaches such as Dirichlet Process mixtures and their extensions have been extensively developed in fields including biostatistics, network analysis, and machine learning.

However, the development of BNP models has been concentrated on statistical and computational domains. Evidence from a systematic review of 43 Scopus-indexed peer-reviewed articles indicates a strong methodological orientation, reflected in the concentration of publications in statistical journals and the predominance of technical concepts related to mixture models, clustering, and Bayesian inference.

In contrast, their application in marketing, particularly in market segmentation and consumer behavior, remains limited. Most segmentation studies rely on traditional clustering or parametric approaches, which require a priori specification of the number of segments and may fail to capture emergent and dynamic consumer heterogeneity. This reveals a gap between methodological advances and their adoption in applied marketing contexts.

This gap is increasingly relevant given the complexity of consumer behavior and the availability of high-dimensional and survey-based data, which require flexible, data-driven approaches capable of identifying latent structures without restrictive assumptions. This study aims to systematically review the literature on BNP models for clustering, with a focus on their applicability in market segmentation. Using a PRISMA-based approach,

Conciencia Digital

ISSN: 2600-5859
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original

www.concienciadigital.org

the study integrates bibliometric and thematic analyses to examine research trends and methodological developments. Additionally, taxonomy is proposed to structure the literature into six categories, ranging from foundational mixture models to advanced hierarchical and data-adaptive extensions.
The study is guided by the following research questions:
RQ1. What evidence exists in Scopus-indexed, peer-reviewed literature regarding the use of Bayesian nonparametric models for clustering and segmentation, and what are their predominant methodological characteristics?
RQ2. What methodological approaches have been proposed using dynamic Bayesian nonparametric models for market segmentation, and how do they capture the emergence and temporal evolution of consumer segments?
RQ3. What studies have applied Bayesian nonparametric models to youth or student market segmentation, and what are their methodological approaches and application contexts?
RQ4. What studies have employed Bayesian nonparametric models for segmentation based on survey data with ordinal or attitudinal variables?
The remainder of this paper is organized as follows. Section two describes the PRISMA-based methodology. Section three presents bibliometric analysis. Section four develops the proposed taxonomy and discusses the findings and research implications. Section five concludes the study and outlines directions for future research.

2. Methodology

This study adopts a systematic literature review approach based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework to ensure transparency, reproducibility, and methodological rigor in the selection and analysis of the literature.
The literature search was conducted in the Scopus database on February 22, 2026. Four structured search queries were designed in alignment with the research questions, combining terms related to Bayesian nonparametric models (e.g., "Dirichlet process mixture," "hierarchical Dirichlet process," "Bayesian nonparametric") with keywords associated with clustering, segmentation, and consumer behavior.
The results from the four search strategies were combined into a unified dataset prior to screening. The initial search yielded 1,121 records. Subsequently, inclusion criteria were applied, including publication year, subject areas (social sciences, business, decision sciences, and psychology), document type (peer-reviewed articles), publication stage (final), language (English), and accessibility. This filtering process reduced the sample to 64 articles.
Duplicate records were then removed, resulting in 54 articles. A final screening phase was conducted based on relevance to the research objectives, leading to a final sample of

ISSN: **2600-5859**
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original
www.concienciadigital.org

48 articles. Studies that did not explicitly address clustering, latent heterogeneity, or Bayesian nonparametric modeling were excluded.

The selection process ensured that only peer-reviewed studies with explicit methodological contributions to Bayesian nonparametric clustering or segmentation were retained. The final dataset constitutes the basis for subsequent bibliometric and thematic analyses.

This study does not involve human participants, personal data, or experimental interventions. Therefore, it is not subject to ethical approval under the principles of the Declaration of Helsinki. Nevertheless, the research was conducted following the ethical standards for scientific publication, ensuring transparency, accuracy, and integrity in accordance with the guidelines of the Committee on Publication Ethics (COPE).

3. Results

This section presents the empirical findings derived from the systematic review. The results include a bibliometric analysis of the selected studies and a thematic classification based on their methodological characteristics. Together, these analyses provide a structured overview of the evolution and organization of the field.

Table 1

Selected scientific articles

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
1	Gwee et al. (2025) Model-based clustering for network data via a latent shrinkage position cluster model. Network Science, 13, e19, 1–33. https://doi.org/10.1017/nws.2025.10014	Propose a Bayesian nonparametric latent shrinkage position cluster model (LSPCM) for model-based clustering of network data, enabling automatic inference of the number of clusters and latent dimensions.	Introduces a Bayesian nonparametric clustering model that combines overfitted finite mixtures with shrinkage priors, allowing automatic determination of the number of clusters and uncertainty quantification.	No.	No.	Network data.

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
2	Giampino et al. (2025) A Bayesian model for co-clustering ordinal data with informative missing entries. Statistics and Computing, 35, 164. https://doi.org/10.1007/s11222-025-10703-w	Develop a Bayesian nonparametric co-clustering model for ordinal data with informative missing values, allowing simultaneous inference of row and column clusters without pre-specifying their number.	Proposes a Bayesian co-clustering framework based on independent Dirichlet processes, latent continuous variables, Bayesian matrix factorization and Gibbs sampling for flexible clustering of ordinal data.	No	No	Ordinal data with informative missing values
3	Battiston & Rlmella (2025) Disclosure risk assessment with Bayesian non-parametric hierarchical modelling. Statistics and Computing, 35, 158. https://doi.org/10.1007/s11222-025-10693-9	Develop a Bayesian nonparametric hierarchical mixed-membership model based on the Hierarchical Dirichlet Process for modeling latent profiles and assessing disclosure risk.	Extends Bayesian mixed-membership models through a Hierarchical Dirichlet Process, enabling an unbounded number of latent components, hierarchical probabilistic modeling, and MCMC inference for heterogeneous data.	Yes (HDP)	No	Categorical profile data

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
4	Park et al. (2025) Nonparametric Bayesian Poisson hurdle random effects model: An application to temperature–suicide association study. Environmental and Ecological Statistics, 32, 579–601. https://doi.org/10.1007/s10651-025-00655-9	Develop a Bayesian nonparametric Poisson hurdle random-effects model based on Dirichlet Process mixtures to analyze zero-inflated count data with structural heterogeneity.	Introduces a Bayesian nonparametric hierarchical model that combines Poisson hurdle regression, Dirichlet Process mixtures, random effects, and MCMC inference to identify latent subgroups in heterogeneous count data.	Yes (Hierarchical DP mixture model)	No	Zero-inflated count data
5	Verde & Rosner (2025) A bias-corrected Bayesian nonparametric model for combining studies with varying quality in meta-analysis. Biometrical Journal, 67, e70034. https://doi.org/10.1002/bimj.70034	Develop a bias-corrected Bayesian nonparametric model based on Dirichlet Process mixtures to combine studies with varying methodological quality in meta-analysis.	Introduces a Bayesian hierarchical framework that integrates Dirichlet Process mixtures with bias correction, allowing flexible modeling of between-study heterogeneity through MCMC inference.	Yes (Hierarchical DP mixture model)	No	Meta-analysis data (study-level effect sizes)

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
6	Chen et al. (2025) Bayesian subtyping for multi-state brain functional connectome with application on preadolescent brain cognition. Biostatistics, 26(1), kxae045. https://doi.org/10.1093/biostatistics/kxae045	Develop a Bayesian nonparametric clustering model for multi-state brain functional connectivity to identify latent subtypes without pre-specifying the number of clusters.	Integrates Dirichlet Process Mixtures, stochastic block models, feature selection, and variational inference within a hierarchical Bayesian framework for clustering network data.	Yes (Hierarchical DP mixture model)	No	Multi-state brain network data
7	Selva et al. (2025) Pyrichlet: a Python package for density estimation and clustering using Gaussian mixture models. Journal of Statistical Software, 112(8). https://doi.org/10.18637/jss.v112.i08	Develop an open-source Python package for Bayesian nonparametric density estimation and clustering based on infinite mixture models.	Provides computational implementations of Dirichlet Process, Pitman–Yor and Geometric Process mixture models, together with Bayesian inference algorithms for clustering and density estimation.	No	No	General continuous data (Gaussian mixture models)

Conciencia Digital
ISSN: 2600-5859
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original
www.concienciadigital.org

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
8	Benlakhdar et al. (2025) Density-adaptive clustering of multivariate angular data using Dirichlet process mixture models with circular normal distribution for artificial intelligence applications. Statistics, Optimization and Information Computing, 13, 1404–1412. https://doi.org/10.19139/soic-2310-5070-2146	Develop a Bayesian nonparametric clustering model for multivariate angular data using Dirichlet Process Mixtures with von Mises distributions.	Introduces a density-adaptive clustering framework based on Dirichlet Process Mixtures and variational inference for automatically estimating the number of clusters.	No	No	Multivariate angular data
9	Arima et al. (2024) A Bayesian nonparametric approach to correct for underreporting in count data. Biostatistics, 25(3), 904–918. https://doi.org/10.1093/biostatistics/kxad027	Develop a Bayesian nonparametric framework to correct for underreporting in count data through latent clustering of reporting probabilities.	Introduces a Dependent Dirichlet Process implemented through a Probit Stick-Breaking Process, integrating hierarchical Bayesian modeling, spatial effects, and MCMC inference.	Yes (Dependent Dirichlet Process)	No	Underreported count data
10	Niu et al. (2024) Enhancing healthcare decision support through explainable AI models for risk prediction. Decision Support Systems. https://doi.org/10.1016/j.dss.2024.114061	Develop an explainable artificial intelligence model integrating Bayesian nonparametric clustering for disease risk prediction from longitudinal electronic health records.	Combines Dirichlet Process Mixture Models with deep learning, attention mechanisms, and variational Bayesian inference to infer latent clusters automatically.	No	No	Longitudinal electronic health records

Conciencia Digital
ISSN: 2600-5859
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original
www.concienciadigital.org

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
11	Stephenson et al. (2024) Identifying dietary consumption patterns from survey data: A Bayesian nonparametric latent class model. Journal of the Royal Statistical Society: Series B. https://doi.org/10.1093/jrsssb/qnad135	Develop a Bayesian nonparametric latent class model that incorporates survey weights to identify dietary consumption patterns from complex survey data.	Introduces a weighted overfitted latent class model with Dirichlet priors and MCMC inference, allowing the number of latent classes to emerge from the data while accounting for complex survey designs.	Overfitted latent class model (DP-equivalent)	No	Categorical survey data
12	Dew et al. (2024) Detecting routines: Applications to ridesharing customer relationship management. Journal of Marketing Research, 61(2), 368–392. https://doi.org/10.1177/00222437231189185	Develop a Bayesian nonparametric framework to identify temporal behavioral routines from ridesharing usage data for customer relationship management.	Combines inhomogeneous Poisson processes with Bayesian Gaussian Processes to model latent routine and non-routine behaviors using fully Bayesian inference (NUTS).	Gaussian Process temporal model	Customer relationship management	Longitudinal temporal event data
13	Franzolini & Rebaudo (2024) Entropy regularization in probabilistic clustering. Statistical Methods & Applications, 33, 37–60. https://doi.org/10.1007/s10260-023-00716-y	Develop an entropy-regularized procedure to improve partition estimation in probabilistic clustering models.	Introduces entropy regularization as a post-processing strategy for Bayesian clustering, improving partition stability without modifying the underlying probabilistic model.	No	No	General clustering data

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
14	Wang et al. (2024) Spatiotemporal clustering with Neyman–Scott processes via connections to Bayesian nonparametric mixture models. Journal of the American Statistical Association, 119(547), 2382–2395. https://doi.org/10.1080/01621459.2023.2257896	Develop a Bayesian framework for spatiotemporal clustering by linking Neyman–Scott processes with Bayesian nonparametric mixture models.	Establishes the theoretical connection between Neyman–Scott processes, mixture of finite mixtures and Dirichlet Process Mixture Models, together with a collapsed Gibbs sampling algorithm for inference.	Spatiotemporal Bayesian mixture model	No	Spatiotemporal point-process data
15	Lavigne & Liverani (2024) Quantifying the uncertainty of partitions for infinite mixture models. Statistics and Probability Letters, 204, 109930. https://doi.org/10.1016/j.spl.2023.109930	Develop a Bayesian framework to quantify partition uncertainty in Dirichlet Process Mixture Models.	Introduces a methodology to quantify posterior uncertainty of cluster assignments by estimating posterior membership probabilities from MCMC samples.	Dirichlet Process Mixture Model (posterior uncertainty quantification)	Not addressed	General clustering data
16	Goeken et al. (2024) Multimodal preference heterogeneity in choice-based conjoint analysis: A simulation study. Journal of Business Economics, 94, 137–185. https://doi.org/10.1007/s11573-023-01156-6	Evaluate Bayesian models for capturing multimodal preference heterogeneity in choice-based conjoint analysis.	Introduces a Dirichlet Process Mixture Multinomial Logit model to infer latent preference segments without pre-specifying the number of clusters.	Dirichlet Process Mixture Multinomial Logit (DPM-MNL)	Choice-based conjoint analysis / Consumer preference	Choice experiment (CBC) data

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
17	Kim et al. (2023) Dirichlet Process Log Skew-Normal Mixture with a Missing-at-Random-Covariate in Insurance Claim Analysis. Econometrics, 11, 24. https://doi.org/10.3390/econometrics11040024	Develop a Bayesian nonparametric model for insurance claim analysis using Dirichlet Process mixtures with missing covariates.	Introduces an infinite mixture of log skew-normal distributions with Bayesian hierarchical inference for modeling heterogeneous insurance losses.	Dirichlet Process Log Skew-Normal Mixture	Not addressed	Insurance claim data
18	Dijkstra et al. (2023) Identifying temporal trajectories in children's learning data using a Dirichlet process Gaussian process mixture model. Computers in Human Behavior, 145, 107754. https://doi.org/10.1016/j.chb.2023.107754	Identify latent temporal learning trajectories using a Dirichlet Process Gaussian Process mixture model.	Combines Dirichlet Process priors with Gaussian Processes to infer latent trajectory clusters from longitudinal learning data.	Dirichlet Process Gaussian Process Mixture Model (DPGP)	Not addressed	Longitudinal learning trajectories
19	Ren et al. (2023) Bayesian nonparametric mixtures of Exponential Random Graph Models for ensembles of networks. Social Networks, 74, 156–165. https://doi.org/10.1016/j.socnet.2023.03.005	Develop a Bayesian nonparametric clustering framework for ensembles of networks using mixtures of Exponential Random Graph Models.	Introduces a Dirichlet Process Mixture of ERGM models with hierarchical Bayesian inference to identify latent network clusters.	Dirichlet Process Mixture of Exponential Random Graph Models (DP-ERGM)	Not addressed	Network ensemble data

Conciencia Digital
ISSN: 2600-5859
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original
www.concienciadigital.org

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
20	Ferrari (2023) hdpGLM: An R package to estimate heterogeneous effects in generalized linear models using hierarchical Dirichlet process. Journal of Statistical Software, 107(10). https://doi.org/10.18637/jss.v107.i10	Develop an R package for estimating heterogeneous effects in generalized linear models using Bayesian nonparametric clustering.	Introduces a Bayesian regression framework that jointly models latent heterogeneity and regression effects through Dirichlet Process priors, integrating clustering and regression within a unified model.	Hierarchical Dirichlet Process Generalized Linear Model (HDP-GLM)	Not addressed	Generalized linear model data
21	Canale et al. (2022) Importance conditional sampling for Pitman–Yor mixtures. Statistics and Computing, 32, 40. https://doi.org/10.1007/s11222-022-10096-0	Develop an efficient Bayesian inference algorithm for Pitman–Yor mixture models used in density estimation and clustering.	Introduces the Importance Conditional Sampling (ICS) algorithm, improving posterior inference for Bayesian nonparametric mixture models.	Pitman–Yor Process Mixture Model (ICS algorithm)	Not addressed	General mixture model data
22	Sadaghiani & Nunez-Yanez (2025) n-HDP-GNN: Community-aware Bayesian clustering for over-smoothing-resilient, communication-efficient distributed graph neural networks. The Journal of Supercomputing, 81, 1560. https://doi.org/10.1007/s11227-025-08017-9	Develop a Bayesian graph neural network that integrates hierarchical Bayesian clustering to identify latent community structures.	Integrates nested Hierarchical Dirichlet Processes with Graph Neural Networks to learn hierarchical latent communities while improving graph representation learning.	Nested Hierarchical Dirichlet Process (n-HDP)	Not addressed	Graph network data

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
23	Khoi et al. (2025) Dirichlet mixed process integrated Bayesian estimation for individual securities. Journal of Risk and Financial Management, 18(6), 304. https://doi.org/10.3390/jrfm18060304	Develop a Bayesian nonparametric framework for estimating the distribution of individual stock returns using Dirichlet Process Mixtures.	Applies Dirichlet Process Mixture models to flexibly estimate complex financial return distributions through Bayesian MCMC inference.	Dirichlet Process Mixture Model (DPM)	Not addressed	Financial return data
24	Rodríguez-Álvarez et al. (2025) Density regression via Dirichlet process mixtures of normal structured additive regression models. Statistics and Computing, 35, 67. https://doi.org/10.1007/s11222-025-10567-0	Develop a Bayesian nonparametric density regression model based on dependent Dirichlet Process Mixtures for conditional density estimation.	Introduces a flexible density regression framework combining structured additive regression with Dependent Dirichlet Process mixtures.	Dependent Dirichlet Process Mixture (DDP)	Not addressed	Continuous response data with covariates
25	Schiavon (2025) Addressing topic modelling via reduced latent space clustering. Statistical Methods & Applications. https://doi.org/10.1007/s10260-025-00779-z	Develop a Bayesian latent space clustering approach for topic modeling in document collections.	Combines Bayesian latent factor models with clustering in a reduced latent space to identify thematic structures in textual data.	Reduced Latent Space Clustering (RLSC)	Not addressed	Text document data

Conciencia Digital
ISSN: 2600-5859
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original
www.concienciadigital.org

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
26	Ren et al. (2023) Bayesian nonparametric mixtures of exponential random graph models for ensembles of networks. Social Networks, 60, 126–138. https://doi.org/10.1016/j.socnet.2019.05.003	Develop a Bayesian nonparametric clustering framework for ensembles of networks using mixtures of exponential random graph models.	Introduces a Bayesian nonparametric mixture model that combines ERGM with Dirichlet Process priors to infer latent network clusters without predefining their number.	Dirichlet Process Mixture of Exponential Random Graph Models (DPM-ERGM)	Not addressed	Network ensemble data
27	Waters et al. (2015) A Bayesian nonparametric factor model for clustering of multivariate categorical data. Journal of Statistical Planning and Inference, 149, 47–63. https://doi.org/10.1016/j.jspi.2014.07.004	Develop a Bayesian nonparametric latent factor model for clustering multivariate categorical data.	Combines latent factor analysis with Dirichlet Process Mixtures to identify latent clusters and model heterogeneity in categorical data.	Dirichlet Process Latent Factor Model	Not addressed	Multivariate categorical data
28	Sukthong & Charuengramporn (2026) Performance evaluation of imputation techniques for telecommunications customer clustering. ECTI Transactions on Computer and Information Technology, 19(1), 174–192. https://doi.org/10.37936/ecti-cit.2026191.2467	Evaluate the impact of missing data imputation techniques on customer clustering performance in telecommunications.	Compares data imputation strategies and their influence on clustering quality for customer segmentation.	K-means, Hierarchical Clustering and DBSCAN	Customer segmentation (telecommunications)	Telecommunications customer data

Conciencia Digital
ISSN: 2600-5859
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original
www.concienciadigital.org

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
29	Grigorova et al. (2025) An automated machine learning framework for customer segmentation integrating RFM analysis and clustering techniques. International Journal of Financial Studies, 13(2), 243. https://doi.org/10.3390/ijfs13040243	Develop an automated framework for customer segmentation integrating RFM analysis with machine learning clustering techniques.	Integrates RFM analysis, clustering algorithms, and automated validation metrics to improve customer segmentation.	K-means + Hierarchical clustering + Hungarian algorithm	Customer segmentation (financial services)	Customer transactional (RFM) data
30	Karulkar et al. (2025) The growing complexity of consumer choices: Unravelling consumer patterns with K-means and fuzzy logic. Journal of Statistical Theory and Applications, 24, 1165–1195. https://doi.org/10.1007/s44199-025-00143-w	Identify consumer behavior patterns in e-commerce using unsupervised clustering techniques.	Apply classical clustering algorithms to reveal heterogeneous consumer behavior patterns in digital markets.	K-means + Fuzzy C-Means	Consumer segmentation (e-commerce)	E-commerce transaction data
31	Suh (2025) Discovering customer segments through interaction behaviors for home appliance business. Journal of Big Data, 12, 57. https://doi.org/10.1186/s40537-025-01111-y	Identify customer segments from Customer Experience Journey data using interaction-based behavioral analysis.	Combines interaction-based customer analytics with unsupervised learning techniques to identify heterogeneous customer profiles.	Hierarchical clustering + Non-negative Matrix Factorization (NMF)	Customer segmentation (customer experience)	Customer interaction (CEJ/RFV) data
32	Abella et al. (2025) Exploring the spatial segmentation of housing markets from online listings. EPJ Data Science. https://doi.org/10.1140/epjds/s13688-025-00551-z	Identify spatial housing market segments using relational information from online real estate listings.	Develops a network-based framework for identifying emerging spatial market segments through relational data analysis.	Multipartite network analysis + Community detection	Market segmentation (housing market)	Online housing listing data

Conciencia Digital
ISSN: 2600-5859
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original
www.concienciadigital.org

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
33	Wu & Liu (2025) HC-means clustering algorithm for precision marketing on e-commerce platforms. Systems and Soft Computing. https://doi.org/10.1016/j.sasc.2025.200236	Develop an improved clustering algorithm for precision marketing in e-commerce platforms.	Proposes an enhanced clustering algorithm that improves centroid initialization and clustering stability for customer segmentation.	HC-means (improved K-means)	Customer segmentation (e-commerce)	E-commerce customer data
34	González & Kolyada (2025) An analysis of kitchen manufacturers in the Curaçao market by applying Kolyada's business modeling and strategic planning methodology. BioProducts Business. https://doi.org/10.22382/bpb-2025-003	Develop a strategic market segmentation framework for kitchen manufacturers using business modeling and clustering techniques.	Integrates strategic market analysis with clustering methods to identify competitive market niches and consumer profiles.	MVC-1 market clustering methodology	Strategic market segmentation	Market and psychographic data
35	Johan (2025) Implementation of customer segmentation model using K-Means and DBSCAN for fashion industry product transaction. International Journal of Informatics and Information Systems, 7(2), 1–10. http://dx.doi.org/10.62527/joiv.9.6.2978	Identify customer segments in the fashion industry using transaction data and clustering techniques.	Compares K-means and DBSCAN for customer segmentation and evaluates clustering quality using internal validation indices.	K-means + DBSCAN	Customer segmentation (fashion retail)	Customer transaction data
36	Stylianou & Pantelidou (2025) Customer behavior analytics and segmentation in online retail using machine learning techniques. Decision Analytics Journal. https://doi.org/10.1016/j.dajour.2025.100600	Analyze customer behavior and identify consumer segments in online retail using machine learning techniques.	Applies unsupervised machine learning to identify behavioral customer profiles for personalized marketing strategies.	K-means clustering (Elbow method)	Customer segmentation (online retail)	Retail transaction data

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
37	Marín (2025) A fuzzy-XAI framework for customer segmentation and risk detection: Integrating RFM, 2-tuple modeling, and strategic scoring. Mathematics, 13, 2141. https://doi.org/10.3390/math13132141	Develop an explainable framework for customer segmentation and churn risk detection integrating fuzzy clustering and XAI techniques.	Combines fuzzy clustering, linguistic modeling, and explainable AI to improve customer segmentation and strategic decision-making.	Fuzzy C-Means + XAI (SHAP/ LIME)	Customer segmentation and churn analysis	Customer RFM data
38	Maruotti et al. (2019) Assessing the influence of marketing activities on customer behaviors: A dynamic clustering approach. METRON, 77(1), 19–42. https://doi.org/10.1007/s40300-019-00150-9	Develop a dynamic clustering framework to analyze the influence of marketing activities on customer behavior over time.	Introduces a dynamic clustering model based on Hidden Markov Models to capture temporal evolution of customer segments.	Hidden Markov Model mixture	Customer behavior / Dynamic market segmentation	Longitudinal customer behavioral data
39	Frühwirth-Schnatter et al. (2018) Analysing plant closure effects using time-varying mixture-of-experts Markov chain clustering. Journal of the Royal Statistical Society. https://doi.org/10.1214/17-AOAS1132	Develop a Bayesian dynamic clustering model for identifying latent groups in longitudinal processes.	Introduces a time-varying mixture-of-experts Markov chain model that captures heterogeneous dynamic trajectories.	Time-varying Mixture-of-Experts Markov Chain	Not addressed	Longitudinal categorical process data

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
40	Linero (2023) Prior and posterior checking of implicit causal assumptions. Biometrics, 79(4), 3153–3164. https://doi.org/10.1111/biom.13886	Develop Bayesian nonparametric tools for modeling latent structures and evaluating implicit modeling assumptions.	Demonstrates the flexibility of Bayesian nonparametric mixture models based on Dirichlet Processes for representing latent heterogeneity through MCMC inference.	Dirichlet Process Mixture Model	Not addressed	General latent structure data
41	Diana et al. (2023) A general modeling framework for open wildlife populations based on the Polya tree prior. Biometrics, 79(3), 2171–2183. https://doi.org/10.1111/biom.13756	Develop a Bayesian nonparametric framework for modeling open wildlife populations using Polya Tree priors.	Uses Polya Tree priors to flexibly model unknown densities associated with entry, residence and exit processes in open populations.	Polya Tree prior model	Not addressed	Ecological population data
42	Yu & Quinn (2022) A multidimensional pairwise comparison model for heterogeneous perceptions. Journal of the Royal Statistical Society: Series A, 185(3), 1049–1073. https://doi.org/10.1111/rssa.12810	Develop a multidimensional pairwise comparison model to capture heterogeneous perceptions among individuals.	Integrates Dirichlet Process Mixture modeling to identify latent groups of evaluators with similar perception patterns.	Dirichlet Process Mixture Model	Not addressed	Pairwise perception data

Table 1

Selected scientific articles (continuation)

No.	Reference	Study objective	Methodological contribution (RQ1)	Methodological extension (RQ2)	Marketing application (RQ3)	Data type (RQ4)
43	Janicki et al. (2022) Bayesian nonparametric multivariate spatial mixture mixed effects models with application to American Community Survey special tabulations. The Annals of Applied Statistics, 16(1), 144–168. https://doi.org/10.1214/21-AOAS1494	Develop a Bayesian nonparametric multivariate spatial mixture mixed-effects model for small area estimation using complex survey data.	Extends spatial mixed-effects modeling with Dirichlet Process mixtures to capture latent multivariate and spatial heterogeneity.	Spatial Dirichlet Process mixture mixed-effects model	Not addressed	Complex survey / small area estimation data

3.1. Bibliometrics

The bibliometric analysis was conducted to provide a structured overview of the scientific production related to Bayesian nonparametric models for clustering and segmentation. The analysis is based on a corpus of 43 peer-reviewed articles, extracted and validated through the PRISMA methodology.

The dataset comprises a total of 133 contributing authors, indicating a dispersed authorship structure, with an average number of authors per paper suggesting a predominance of collaborative research. In terms of institutional diversity, the studies are affiliated with 90 different institutions across 15 countries, reflecting a moderate level of internationalization in the field. However, the collaboration structure suggests that research is still fragmented rather than dominated by consolidated research networks.

Regarding publication sources, the articles are distributed across 32 different journals, which indicates a heterogeneous dissemination of knowledge rather than concentration in a small number of outlets. This dispersion is consistent with an emerging research field that intersects multiple disciplines, including statistics, data science, and applied analytics.

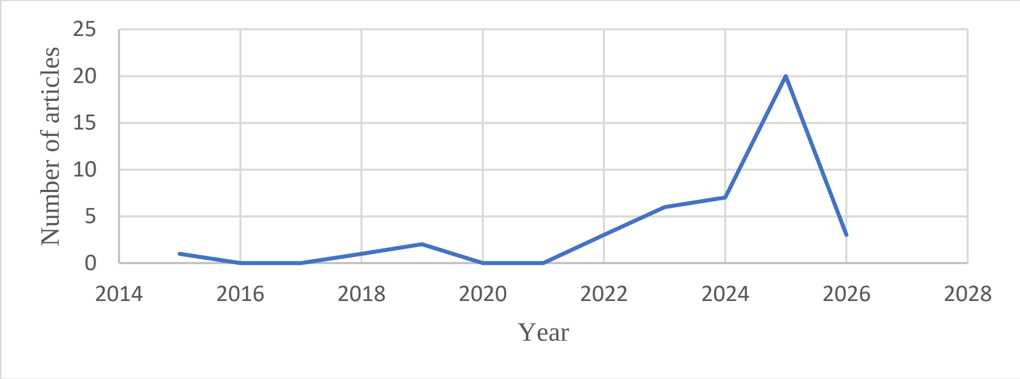
In terms of scientific impact, the dataset includes a set of highly cited papers alongside a broader distribution of moderately cited studies, suggesting a combination of foundational contributions and recent developments. Additionally, the keyword structure is extensive, with 196 author keywords and 173 indexed keywords, reflecting a strong diversity of research topics, although oriented toward methodological developments.

Overall, these indicators suggest that the field of Bayesian nonparametric clustering is characterized by methodological maturity but remains fragmented in terms of application domains, particularly in marketing and consumer research. The following sections present a detailed analysis of scientific production, sources, and thematic structure.

Conciencia Digital
ISSN: 2600-5859
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original
www.concienciadigital.org

Figure 1

Annual scientific production



Note: the figure shows the number of publications per year based on the articles included in the systematic review. The apparent decrease in 2026 reflects the partial coverage of the current year.

To further explore the structure of the field, the following figures present the temporal evolution of publications, the main publication sources, and the thematic distribution of research topics.

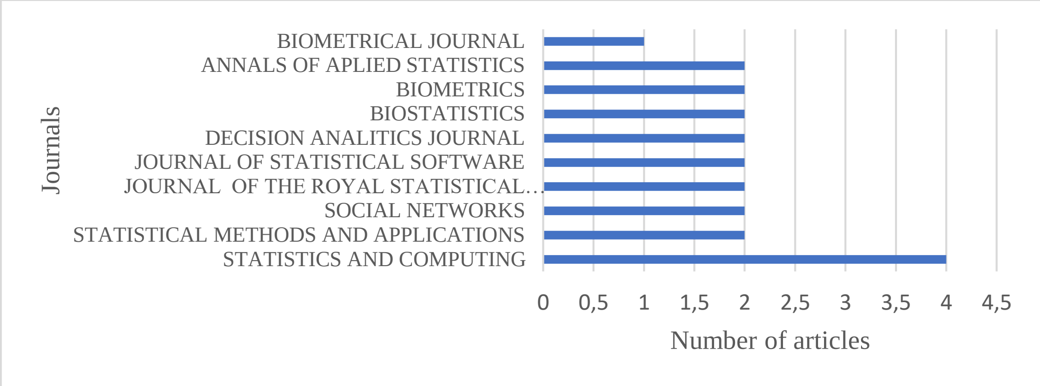
Figure 1 presents the annual scientific production of studies included in the review. The results show a clear upward trend, particularly from 2022 onwards, indicating a growing academic interest in Bayesian nonparametric approaches.

While early contributions were sporadic between 2015 and 2019, a more consistent increase is observed from 2022, followed by a sharp rise in 2025, which represents the peak in the number of publications. This pattern suggests that the field is rapidly expanding, driven by the increasing need to model complex and dynamic data structures. The apparent decline in 2026 should be interpreted with caution, as it reflects the partial coverage of the current year rather than a decrease in research activity. This growth further supports the relevance of systematically analyzing the development of Bayesian nonparametric models and their potential applications in market segmentation.

Figure 2 presents the distribution of publications across the most relevant journals, providing an overview of the main outlets in which research on Bayesian nonparametric models has been published.

Figure 2

Top journals publishing Bayesian nonparametric research



Note: number of articles per journal based on the studies included in the bibliometric analysis.

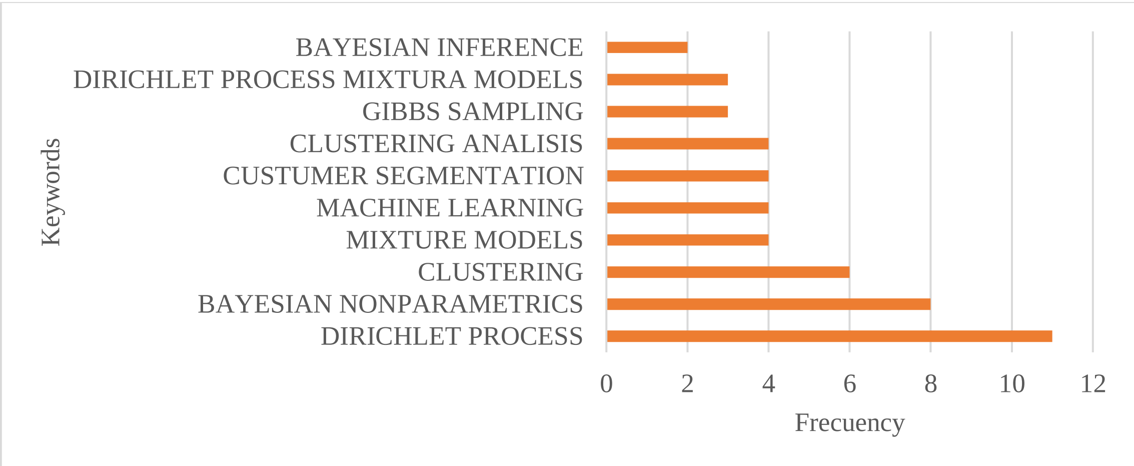
The distribution of publications across journals shows a clear concentration in statistical and methodological outlets, with Statistics and Computing standing out as the leading source. Other prominent journals, such as Statistical Methods and Applications, Annals of Applied Statistics, Biometrics, and Biostatistics, further confirm the strong methodological orientation of the field.

In contrast, marketing-oriented journals are notably absent among the most relevant sources. This pattern indicates that, despite the potential of Bayesian nonparametric models for segmentation, their application in market segmentation and consumer behavior research remains limited. Overall, these findings highlight a gap between methodological developments and their adoption in marketing contexts.

Figure 3 presents the most frequent author keywords, providing an overview of the main research topics addressed in the field.

Figure 3

Most frequent author keywords



Note: The vertical axis lists the author keywords, while the horizontal axis shows their frequency of occurrence across the studies included in the bibliometric analysis.

The keyword analysis reveals a strong concentration in methodological concepts, with "Dirichlet process" and "Bayesian nonparametric" emerging as the most frequent terms. Other prominent keywords, such as "clustering," "mixture models," and "Gibbs sampling," further reinforce the technical and model-oriented nature of the field.

Although "customer segmentation" appears among the most frequent keywords, its relative presence remains limited compared to methodological terms. This pattern suggests that, while the potential application of Bayesian nonparametric models in market segmentation is recognized, their use in marketing contexts is still not widely developed.

Overall, these findings highlight a clear emphasis on methodological advancements, with comparatively less focus on applied research in consumer behavior and marketing, reinforcing the gap identified between theoretical development and practical application.

3.2. Thematic analysis: taxonomy of Bayesian nonparametric models

Based on the analysis of the selected studies, a taxonomy of Bayesian nonparametric models for clustering and segmentation is proposed. Rather than being structured around application domains, the literature is primarily organized according to methodological developments. The reviewed studies can be grouped into six main categories: (1) core nonparametric mixture models, (2) hierarchical and dependent structures, (3) extensions for complex data, (4) inference and computational methods, (5) applied clustering contexts, and (6) market segmentation approaches.

3.2.1. Core nonparametric mixture models

The foundational layer of the literature is composed of Bayesian nonparametric mixture models, particularly Dirichlet Process Mixtures (DPM) and related infinite mixture formulations. These models allow the number of clusters to be inferred directly from the data, avoiding the need for predefined model complexity (Ferguson, 1973; Teh et al., 2006).

Several studies employ these approaches for density estimation and clustering, highlighting their flexibility in capturing latent heterogeneity across different domains (Stephenson et al., 2024; Dew et al., 2024). The use of stick-breaking representations and fully Bayesian inference frameworks enable a probabilistic interpretation of clustering structures.

Despite their strong theoretical foundation, these models are developed within statistical and computational research contexts rather than applied marketing settings.

3.2.2. Hierarchical and dependent structures

A second group of studies extends these models through hierarchical and dependency-based frameworks, such as the Hierarchical Dirichlet Process (HDP) and related dependent Bayesian nonparametric models (Teh et al., 2006). Recent applications have expanded these approaches to accommodate multilevel, temporal, spatial and structured dependencies in complex datasets (Ferrari, 2023; Dijkstra et al., 2023; Wang et al., 2024; Giampino et al., 2025; Sadaghiani & Nunez-Yanez, 2025). These approaches allow for the modeling of structured heterogeneity, incorporating temporal, spatial, or multilevel dependencies.

In these models, clustering is no longer static but can evolve according to temporal, spatial or contextual dependencies, enabling the identification of condition-dependent groups and more complex latent structures (Dijkstra et al., 2023; Wang et al., 2024; Park et al., 2025; Sadaghiani & Nunez-Yanez, 2025). This added flexibility enhances the capacity of Bayesian nonparametric methods to represent complex data structures. However, their application remains concentrated in methodological and domain-specific studies outside marketing.

3.2.3. Extensions for complex data structures

The literature also reports on a wide range of methodological extensions designed to address specific data structures, including ordinal variables, network data, temporal processes, and high-dimensional observations. These developments comprise Bayesian co-clustering models, Gaussian process mixture models, and network-based clustering approaches, thereby expanding the applicability of Bayesian nonparametric methods to increasingly complex analytical settings (Waters et al. 2015; Dijkstra et al., 2023; Ren et al., 2023; Gwee et al., 2025; Giampino et al., 2025).

These approaches enable the identification of latent structures in complex datasets by integrating domain-specific features into the modeling framework. For example, co-clustering models allow simultaneous clustering of observations and variables, while Gaussian process mixtures capture smooth functional patterns.

While these developments demonstrate the versatility of Bayesian nonparametric methods, their applications remain concentrated in fields such as biostatistics and machine learning.

3.2.4. Inference and computational methods

A massive portion of the literature focuses on inference and computational strategies required to estimate Bayesian nonparametric models. Due to their infinite-dimensional nature, these models require specialized algorithms to approximate posterior distributions.

Common inference strategies identified in the reviewed literature include Markov Chain Monte Carlo (MCMC) algorithms, particularly Gibbs sampling, together with other Bayesian estimation procedures aimed at efficiently approximating posterior distributions. Recent studies have also emphasized the development of computationally efficient implementations that improve scalability and facilitate the application of Bayesian nonparametric models to increasingly complex datasets (Arima et al., 2024; Benlakhdar et al., 2025; Chen et al., 2025; Giampino et al., 2025; Selva et al., 2025)

These methodological contributions are essential for enabling the practical implementation of Bayesian nonparametric models in real-world scenarios.

3.2.5. Applied clustering contexts

Empirical applications identified in the reviewed literature span a wide range of domains, including healthcare, finance, insurance, network analysis, and evidence synthesis. Across these application areas, Bayesian nonparametric models have consistently demonstrated their ability to uncover latent structures, accommodate complex sources of

Conciencia Digital
ISSN: 2600-5859
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original
www.concienciadigital.org

heterogeneity, and provide flexible probabilistic representations of real-world phenomena (Kim et al., 2023; Ren et al., 2023; Arima et al., 2024; Gwee et al., 2025; Khoi et al., 2025; Park et al., 2025; Verde & Rosner, 2025).

Only a limited number of studies apply these approaches in contexts related to customer behavior and segmentation (Dew et al., 2024), highlighting their potential relevance for marketing research. However, these applications remain scarce compared to methodological developments. This imbalance suggests that the practical adoption of Bayesian nonparametric methods in marketing analytics is still at an early stage.

3.2.6. Market segmentation and the methodological Gap

In contrast, the reviewed literature on market segmentation continues to be dominated by conventional clustering techniques, including K-means, hierarchical clustering, density-based methods, and other machine learning approaches. Although these techniques have demonstrated their usefulness in customer segmentation, they generally require the number of clusters to be specified a priori or rely on heuristic optimization procedures, while lacking a fully probabilistic framework for representing uncertainty in cluster allocation (Johan, 2025; Grigorova et al., 2025; Karulkar et al., 2025; Stylianou & Pantelidou, 2025; Suh, 2025; Wu & Liu, 2025).

The comparison between both streams of research reveals a clear methodological gap. While Bayesian nonparametric models provide a flexible and theoretically grounded framework for modeling latent heterogeneity and emergent segmentation, their integration into marketing research remains limited. This gap is particularly relevant in the context of youth markets and survey-based data, where flexible and data-driven segmentation approaches could offer substantial analytical advantages.

4. Discussion

This study analyzed a corpus of 48 peer-reviewed articles, of which 43 were included in the bibliometric analysis, revealing a consistent asymmetry between methodological development and applied research in Bayesian nonparametric clustering. The reviewed literature is oriented towards methodological innovation, focusing on flexible mixture-based frameworks, hierarchical extensions, and advanced inference procedures, particularly those based on Dirichlet processes and related constructions (Gwee et al., 2025; Giampino et al., 2025; Battiston & Rimella, 2025).

These approaches enable the estimation of latent heterogeneity without predefined structures and have been successfully applied in domains such as network analysis, biostatistics, and high-dimensional data modeling (Chen et al., 2025; Ren et al., 2023; Janicki et al., 2022). Additional contributions emphasize improvements in inference and computational efficiency, including entropy-regularized clustering and advanced sampling methods (Franzolini & Rebaudo, 2024; Selva et al., 2025; Canale et al., 2022). Despite these advances, their application in marketing, particularly in market segmentation, remains limited. Only a small subset of studies explores segmentation

Conciencia Digital
ISSN: 2600-5859
Vol. 9 No. 3, pp. 28 – 59, julio - septiembre 2026
Revista multidisciplinar
Articulo original
www.concienciadigital.org

contexts, through extensions of latent class models and Dirichlet Process Mixtures. These include applications to survey-based segmentation (Stephenson et al., 2024), preference heterogeneity in conjoint analysis (Goeken et al., 2024), and temporal customer behavior modeling (Dew et al., 2024), demonstrating the potential of Bayesian nonparametric approaches in marketing settings.

In contrast, the broader segmentation literature continues to rely on traditional clustering and machine learning techniques, such as K-means, fuzzy clustering, and hybrid approaches, which remain dominant in applied contexts (Grigorova et al., 2025; Karulkar et al., 2025; Marín 2025; Suh, 2025). These methods typically require a priori specification of segment structures and provide limited flexibility for modeling emergent patterns.

This contrast highlights a structural gap between methodological advances and their application in marketing research. While Bayesian nonparametric models offer a theoretically robust framework for modeling latent heterogeneity and identifying emergent segments, their adoption remains constrained by factors such as computational complexity, scalability, and limited accessibility of user-friendly implementations (Lavigne & Liverani, 2024; Ferrari, 2023).

From a theoretical perspective, this study contributes by organizing the literature into a taxonomy that reflects the evolution of Bayesian nonparametric clustering models, from foundational mixture models to hierarchical, dependent, and data-adaptive extensions (Arima et al., 2024; Sadaghiani & Nunez-Yanez, 2025).

From a methodological standpoint, the findings confirm the capacity of Bayesian nonparametric approaches to model complex data structures, including ordinal, temporal, and high-dimensional data (Giampino et al., 2025; Dijkstra et al., 2023). From an applied perspective, the results identify a clear opportunity for future research: the integration of Bayesian nonparametric models into market segmentation, particularly in contexts involving survey data, ordinal variables, and dynamic consumer behavior.

5. Conclusions

- This systematic review synthesized the evidence available in Scopus-indexed, peer-reviewed literature on the application of Bayesian nonparametric models to clustering and segmentation.
- Regarding RQ1, the review showed that Bayesian nonparametric models have been extensively developed for clustering and segmentation, through Dirichlet Process Mixture Models and their extensions, including Hierarchical and Dependent Dirichlet Processes. These approaches allow the number of clusters to be inferred from the data and provide a flexible representation of latent heterogeneity.
- Regarding RQ2, the evidence indicates that dynamic approaches remain limited. The identified studies employ Hidden Markov Models and mixture-based

dynamic models rather than fully dynamic Bayesian nonparametric frameworks, indicating that the identification of emerging consumer segments over time has received limited attention.

- Regarding RQ3, no studies were identified that applied Bayesian nonparametric models to the segmentation of youth or university student markets. Existing applications are concentrated in areas such as healthcare, finance, insurance, ecology, network analysis, and general customer analytics, whereas marketing studies employ traditional clustering methods.
- Regarding RQ4, only a small number of studies applied Bayesian nonparametric models to survey data containing ordinal or attitudinal variables. These studies demonstrate the suitability of Bayesian nonparametric approaches for modeling latent heterogeneity in survey-based data; however, their application to market segmentation remains limited.
- Overall, the findings indicate that, despite the substantial methodological development of Bayesian nonparametric models, their application to market segmentation, particularly in youth markets using survey-based demographic, attitudinal, and behavioral data, remains scarce. This gap supports the need for further empirical research integrating Bayesian nonparametric methods into marketing segmentation.

6. Conflict of interest

The authors declare that there is no conflict of interest in relation to the article presented.

7. Authors' contribution statement

All authors contributed significantly to the elaboration of the article.

7.1. Disclosure of delegation to generative AI

The authors declare the use of generative AI in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision:
- Idea generation
- Formulation of research questions and hypotheses
- Search and systematization of literature

The GAI tool used was: ChatGPT 4.
The responsibility for the final manuscript lies entirely with the authors.
GAI tools are not listed as authors and do not assume responsibility for the results.
Statement submitted by: Name of lead author.

https://panbibliotekar.github.io/gaidet-declaration/index.html

8. Financing costs

This research was funded entirely with the authors' own funds.

9. References

Abella, D., Martínez, J. H., Mazzoli, M., Migozzi, J., Le Corre, T., Alonso-Paulí, E., Crespí-Cladera, R., Louail, T., & Ramasco, J. J. (2025). Exploring the spatial segmentation of housing markets from online listings. *EPJ Data Science*, 14(34). https://doi.org/10.1140/epjds/s13688-025-00551-z

Arima, S., Polettini, S., Pasculli, G., Gesualdo, L., Pesce, F., & Procaccini, D.-A. (2024). A Bayesian nonparametric approach to correct for underreporting in count data. *Biostatistics*, 25(3), 904–918. https://doi.org/10.1093/biostatistics/kxad027

Battiston, M., & Rimella, L. (2025). Disclosure risk assessment with Bayesian non-parametric hierarchical modelling. *Statistics and Computing*, 35, 158. https://doi.org/10.1007/s11222-025-10693-9

Benlakhdar, S., Nadarajah, S., Rziza, M., & Oulad Haj Thami, R. (2025). Density-adaptive clustering of multivariate angular data using Dirichlet process mixture models with circular normal distribution for artificial intelligence applications. *Statistics Optimization and Information Computing*, 13(4), 1404–1412. https://doi.org/10.19139/soic-2310-5070-2146

Canale, A., Corradin, R., & Nipoti, B. (2022). Importance conditional sampling for Pitman–Yor mixtures. *Statistics and Computing*, 32, 40. https://doi.org/10.1007/s11222-022-10096-0

Chen, T., Zhao, H., Tan, C., Constable, T., Yip, S., & Zhao, Y. (2025). Bayesian subtyping for multi-state brain functional connectome with application on preadolescent brain cognition. *Biostatistics*, 26(1), kxae045. https://doi.org/10.1093/biostatistics/kxae045

Dew, R., Ascarza, E., Netzer, O., & Sicherman, N. (2024). Detecting routines: Applications to ridesharing customer relationship management. *Journal of Marketing Research*, 61(2), 368–392. https://doi.org/10.1177/00222437231189185

Diana, A., Matechou, E., Griffin, J., Arnold, T., Tenan, S., & Volponi, S. (2023). A general modeling framework for open wildlife populations based on the Polya tree prior. *Biometrics*, *79*(3), 2171–2183. https://doi.org/10.1111/biom.13756

Dijkstra, S. H. E., Hinne, M., Segers, E., & Molenaar, I. (2023). Clustering children's learning behaviour to identify self-regulated learning support needs. *Computers in Human Behavior*, 145 (107754), 107754. https://doi.org/10.1016/j.chb.2023.107754

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2), 209–230. https://doi.org/10.1214/aos/1176342360

Ferrari, D. (2023). hdpGLM: An R package to estimate heterogeneous effects in generalized linear models using hierarchical Dirichlet process. *Journal of Statistical Software*, 107(10), 1-37. https://doi.org/10.18637/jss.v107.i10

Franzolini, B., & Rebaudo, G. (2024). Entropy regularization in probabilistic clustering. *Statistical Methods & Applications*, 33, 37–60. https://doi.org/10.1007/s10260-023-00716-y

Frühwirth-Schnatter, S., Pittner, S., Weber, A., & Winter-Ebmer, R. (2018). Analyzing plant closure effects using time-varying mixture-of-experts Markov chain clustering. *The Annals of Applied Statistics*, 12(3), 1796-1830. https://doi.org/10.1214/17-aoas1132

Giampino, A., Canale, A., & Nipoti, B. (2025). A Bayesian model for co-clustering ordinal data with informative missing entries. *Statistics and Computing*, 35, 164. https://doi.org/10.1007/s11222-025-10703-w

Goeken, N., Kurz, P., & Steiner, W. J. (2024). Multimodal preference heterogeneity in choice-based conjoint analysis: A simulation study. Journal of Business Economics, 94(1), 137–185. https://doi.org/10.1007/s11573-023-01156-6

González Rodríguez, A. J., & Kolyada, A. (2025). An analysis of kitchen manufacturers in the Curaçao market by applying Kolyada's business modeling and strategic planning methodology. *BioProducts Business*, 10(3). https://doi.org/10.22382/bpb-2025-003

Grigorova, I., Efremov, A., & Karamfilov, A. (2025). An automated machine learning framework for customer segmentation integrating RFM analysis and clustering techniques. *International Journal of Financial Studies*, 13(4), 243. https://doi.org/10.3390/ijfs13040243

Gwee, X. Y., Gormley, I. C., & Fop, M. (2025). Model-based clustering for network data via a latent shrinkage position cluster model. *Network Science*, 13(e19). https://doi.org/10.1017/nws.2025.10014

Janicki, R., Raim, A. M., Holan, S. H., & Maples, J. J. (2022). Bayesian nonparametric multivariate spatial mixture mixed effects models with application to American Community Survey special tabulations. *The Annals of Applied Statistics*, 16(1), 144–168. https://doi.org/10.1214/21-AOAS1494

Johan, M. E., (2025). Implementation of customer segmentation model using K-Means and DBSCAN for fashion industry product transaction. *International Journal on Informatics Visualization*, 9(6). http://dx.doi.org/10.62527/joiv.9.6.2978

Karulkar, Y., Srivastava, S., Nandwana, R., & Stanley, S. (2025). The growing complexity of consumer choices: Unravelling consumer patterns with K-means and fuzzy logic. *Journal of Statistical Theory and Applications*, 24, 1165–1195. https://doi.org/10.1007/s44199-025-00143-w

Khoi, P. D., Trong, T. M., & Gan, C. (2025). Dirichlet mixed process integrated Bayesian estimation for individual securities. *Journal of Risk and Financial Management*, 18(6), 304. https://doi.org/10.3390/jrfm18060304

Kim, M., Lindberg, D., Crane, M., & Bezbradica, M. (2023). Dirichlet process log skew-normal mixture with a missing-at-random-covariate in insurance claim analysis. *Econometrics*, 11(4), 24. https://doi.org/10.3390/econometrics11040024

Lavigne, A., & Liverani, S. (2024). Quantifying the uncertainty of partitions for infinite mixture models. *Statistics and Probability Letters*, 204 (109930), 109930. https://doi.org/10.1016/j.spl.2023.109930

Linero, A. R. (2023). Prior and posterior checking of implicit causal assumptions. *Biometrics*, 79(4), 3153–3164. https://doi.org/10.1111/biom.13886

Marín Díaz, G. (2025). A fuzzy-XAI framework for customer segmentation and risk detection: Integrating RFM, 2-tuple modeling, and strategic scoring. *Mathematics*, 13(13), 2141. https://doi.org/10.3390/math13132141

Maruotti, A., Bulla, J., & Mark, T. (2019). Assessing the influence of marketing activities on customer behaviors: a dynamic clustering approach. Metron. *International Journal of Statistics*, 77, 19–42. https://doi.org/10.1007/s40300-019-00150-9

Niu, S., Yin, Q., Maa, J., Song, Y., Xu, Y., Bai, L., Pan, W., Yang, X. (2024). Enhancing healthcare decision support through explainable AI models for risk prediction. *Decision Support Systems*, 181(114228), 114228. https://doi.org/10.1016/j.dss.2024.114228

Park, J., Sim, K., Yang, D., Kim, Y., Hashizume, M., & Chung, Y. (2025). Nonparametric Bayesian Poisson hurdle random effects model: An application to temperature–suicide association study. *Environmental and Ecological Statistics*, 32, 579–601. https://doi.org/10.1007/s10651-025-00655-9

Ren, S., Wang, X., Liu, P., & Zhang, J. (2023). Bayesian nonparametric mixtures of exponential random graph models for ensembles of networks. *Social Networks*, 74, 156–165. https://doi.org/10.1016/j.socnet.2023.03.005

Rodríguez-Álvarez, M. X., Inácio, V., & Klein, N. (2025). Density regression via Dirichlet process mixtures of normal structured additive regression models. *Statistics and Computing*, 35, 47. https://doi.org/10.1007/s11222-025-10567-0

Sadaghiani, A. K., & Nunez-Yanez, J. (2025). n-HDP-GNN: Community-aware Bayesian clustering for distributed graph neural networks. *The Journal of Supercomputing*, 81, 1560. https://doi.org/10.1007/s11227-025-08017-9

Schiavon, L. (2025). Addressing topic modelling via reduced latent space clustering. *Statistical Methods & Application, 34, 1–20.* https://doi.org/10.1007/s10260-025-00779-z

Selva, F., Fuentes-García, R., & Gil-Leyva, M. F. (2025). Pyrichlet: a Python package for density estimation and clustering using Gaussian mixture models. *Journal of Statistical Software*, 112(8), 1–39. https://doi.org/10.18637/jss.v112.i08

Stephenson, B. J. K., Wu, S. M., & Dominici, F. (2024). Identifying dietary consumption patterns from survey data: a Bayesian nonparametric latent class model. *Journal of the Royal Statistical Society: Series A*, 187(2), 496–512. https://doi.org/10.1093/jrsssa/qnad135

Stylianou, T., & Pantelidou, A. (2025). A machine learning approach to consumer behavior in supermarket analytics. *Decision Analytics Journal*, 16(100600). https://doi.org/10.1016/j.dajour.2025.100600

Suh, Y. (2025). Discovering customer segments through interaction behaviors. *Journal of Big Data*, 12, 57. https://doi.org/10.1186/s40537-025-01111-y

Sukthong, P., & Charuengramporn, P. (2026). Performance evaluation of imputation techniques for telecommunications customer clustering. *ECTI Transactions on Computer and Information Technology*, 20(1), 174–192. https://ph01.tci-thaijo.org/index.php/ecticit/article/view/264267

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). Hierarchical Dirichlet Processes. *Journal of the American Statistical Association*, 101(476), 1566–1581. https://doi.org/10.1198/016214506000000302

Verde, P. E., & Rosner, G. L. (2025). A bias-corrected Bayesian nonparametric model for combining studies with varying quality in meta-analysis. *Biometrical Journal*, 67, e70034. https://doi.org/10.1002/bimj.70034

Wang, Y., Degleris, A., Williams, A., & Linderman, S. W. (2024). Spatiotemporal clustering with Neyman–Scott processes via connections to Bayesian nonparametric mixture models. *Journal of the American Statistical Association*, 119(547), 2382–2395. https://doi.org/10.1080/01621459.2023.2257896

Waters, A., Fronczyk, K., Guindani, M., Baraniuk, R. G., & Vannucci, M. (2015). A Bayesian nonparametric approach for the analysis of multiple categorical item responses. *Journal of Statistical Planning and Inference*, 166, 52–66. https://doi.org/10.1016/j.jspi.2014.07.004

Wu, D., & Liu, X. (2025). HC-means clustering algorithm for precision marketing on e-commerce platforms. *Systems and Soft Computing*, 7(200236), 200236. https://doi.org/10.1016/j.sasc.2025.200236

Yu, Q., & Quinn, K. M. (2022). A multidimensional pairwise comparison model for heterogeneous perceptions with an application to modelling the perceived truthfulness of public statements on COVID-19. *Journal of the Royal Statistical Society. Series A, (Statistics in Society*), 185(3), 1049–1073. https://doi.org/10.1111/rssa.12810

El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Conciencia Digital.**

El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Conciencia Digital.**