
Reader Segmentation on Top 5 Online News Media in Indonesia Using K-Means Clustering Algorithm

Tyas Cahya Larasati*, Jerry Heikal
Universitas Bakrie, Indonesia
Email: tyas.cahya@student.bakrie.ac.id*

KEYWORDS

Online Media; Reader Segmentation; K-Means Clustering; Online Value Proposition; Consumer Behavior; Digital Journalism

ABSTRACT

The rapid growth of digital news consumption in Indonesia has not been matched by accurate audience mapping, leading to suboptimal content personalization and monetization. This research aims to identify reader segmentation across the top five online news media outlets in Indonesia Detik.com, Kompas.com, Tribunnews.com, CNN Indonesia, and TVOneNews.com in order to formulate appropriate Online Value Proposition (OVP) strategies. Using a quantitative approach with K-Means Clustering algorithm on behavioral data from 100 respondents, the optimal number of clusters was determined using the Elbow method. Research variables encompass demographics, content preferences, access frequency, reading duration, and device usage. Three primary clusters were identified: Cluster 1 (Working Women Entertainment Reader) dominated by established professional women favoring entertainment video content; Cluster 2 (Executive Millennial National News Interest) male professional "heavy readers" accessing news over 20 times per day; and Cluster 3 (The Lifestyle Trend-Seeker) Gen Z active TikTok users preferring lifestyle content. Cluster 2 provides the greatest economic potential via traffic volume, while Cluster 1 offers economic stability through household purchasing power. Media outlets are recommended to adopt differentiated OVPs including rapid news summary features (News Bites) for professionals and aesthetic visual content for the younger generation.

INTRODUCTION

The rapid development of digital technology over the past two decades has fundamentally transformed how Indonesian society accesses and consumes information. Online media has become one of the primary news sources, displacing conventional media such as print newspapers and television. According to We Are Social and Hootsuite (2024), more than 213 million internet users in Indonesia access various digital platforms daily, approximately 85% of whom access news via online media. Data from Goodstats.id (2024) reveals the ten most frequently used online news portals in Indonesia, with Detik.com leading at nearly 50% of users, followed by Kompas.com at approximately 40% and Tribunnews.com at nearly 30%. The Reuters Institute Digital News Report (2024) confirms that Indonesia is one of the fastest-growing digital news markets. Despite significantly high traffic with monthly visits ranging from 30 to 60 million users (SimilarWeb, 2024) many Indonesian online media organizations still struggle to accurately map their audience characteristics. Content and advertising strategies remain largely generalized (one-size-fits-all), resulting in suboptimal content personalization, less effective ad placement, and ultimately lower reader engagement and platform monetization efficiency (Zhao et al., 2022; Zhou et al., 2020).

Online media is defined as media based on telecommunication and multimedia technology, including computers and the internet, including portals, websites, social media, and online broadcasting platforms (Asep Syamsul, 2018). In terms of content, online journalism follows the same basic principles as conventional journalism conveying news,

opinions, features, and advertisements which distinguishes only the medium of delivery which is digital (Teguh Trianto, 2016). The quality of news is measured through four values: actual, factual, important, and interesting (Romli, 2016). Although site traffic is very high with monthly visits ranging from 30 to 60 million users (SimilarWeb, 2024) many Indonesian online media organizations still struggle to accurately map the characteristics of their audiences. Content and advertising strategies still tend to be one-size-fits-all, resulting in less than optimal content personalization, less effective ad placement, and ultimately lower reader engagement and platform monetization efficiency (Ramkumar, 2025; Vargo et al., 2024).

Market segmentation, defined by Kotler and Keller (2022), is the process of dividing the market into sub-groups of consumers who are well-defined and have similar needs and interests. Suparyanto and Rosad (2015) classify segmentation variables into geographical, demographic, psychographic, behavioral, and combination variables. In the context of digital media, behavioral variables such as access frequency, reading duration, content preferences, and devices used provide the most granular and actionable segmentation data. To identify these segments empirically, this study uses the K-Means Clustering algorithm, which is an unsupervised machine learning algorithm that partitions the dataset into k clusters by minimizing the Within-Cluster Sum of Squares (WCSS). This algorithm iteratively assigns each data point to the nearest centroid and recalculates the centroid until it converges (Kodinariya & Makwana, 2013). Its advantages include computational efficiency for large behavioral datasets as well as high intra-cluster similarity and low inter-cluster similarity. The Elbow method by plotting WCSS on the number of clusters to find the elbow point as the saturation point is used to determine the optimal k value (Hadiwibowo, 2023).

Furthermore, to formulate the right strategy for each segment of readers, this study refers to the concept of Online Value Proposition (OVP), which is a series of specific benefits and features that digital platforms offer to their target segments to differentiate themselves from competitors (Ghandour, 2019). In the media context, OVP translates into differentiated content strategies, user experience features, and monetization models tailored to the unique needs of each reader segment. Previous research by Garcia-Murillo and MacInnes (2021) showed that audience-specific OVPs significantly improved engagement metrics on digital journalism platforms.

The research gap lies in the fact that although K-Means Clustering has been applied to the segmentation of e-commerce and academic behavior (Lestari & Nugroho, 2021; Putra et al., 2022), as well as audience segmentation has been studied in the context of Western digital journalism (Garcia-Murillo & MacInnes, 2021), there has been no research integrating K-Means Clustering and OVP strategy formulation specifically in the top online news media ecosystems in Indonesia. This study therefore addresses three primary research questions: (1) What reader segmentation profiles exist across the top five Indonesian online news media? (2) Which segment provides the greatest economic value? (3) What differentiated OVP strategies should each media adopt to enhance engagement and retention?

Based on the research questions, this study aims to identify and profile reader segmentation in the top five online news media in Indonesia including Detik.com, Kompas.com, Tribunnews.com, CNN Indonesia, and TVOneNews.com using the K-Means Clustering algorithm optimized through the Elbow method, as well as analyzing the segment of readers that provide the highest economic value based on traffic volume, monetization

potential, and influence on Purchase decision. In addition, this study also aims to formulate a differentiated Online Value Proposition (OVP) strategy that can be implemented by each media to increase reader engagement, retention, and monetization efficiency.

This research is expected to provide both theoretical and practical benefits. Theoretically, this study enriches the literature on digital media audience segmentation by integrating the machine learning approach (K-Means) and the previously separate OVP framework, as well as being one of the first studies to apply behavior-based segmentation in the context of Indonesia's rapidly growing online news market. Practically, the results of this study can be a guide for online media management in developing a more personalized content strategy, targeted ad placement, and developing user experience features that suit the unique characteristics of each reader segment. For advertisers, this study provides insight into the audience segments that are most economically valuable, while for academics and researchers below, it provides a methodological framework that can be replicated and developed with other clustering algorithms or a broader sample.

METHOD

Research Design and Object

This study employed a quantitative survey approach. The research object comprises readers of the five most popular online news media in Indonesia in 2024: Detik.com, Kompas.com, Tribunnews.com, TVOneNews.com, and CNNIndonesia.com. These platforms were selected based on their dominant position in Indonesian web traffic rankings (SimilarWeb, 2024), each recording 30–60 million monthly visits.

Sample and Data Collection

The population encompasses online news readers across Indonesia's six major island groups: Sumatra, Java, Kalimantan, Sulawesi, Bali, and Papua. Given the vast population size, a purposive sample of 100 respondents was employed, consistent with established clustering research norms (Sugiyono, 2019). Data were collected via structured Google Form questionnaire distributed between November 2024 and February 2025. The instrument underwent expert judgement validation by three domain specialists before deployment.

Research Variables

Nineteen variables were operationalized across two dimensions. Table 2 presents the key variables used in the K-Means analysis.

Table 1. Key Research Variables (Operationalization)

Variable	Indicator	Scale	Code
Demographics	Gender, Age, Education, Occupation, Marital Status, Domicile	Nominal/Ordinal	X1–X6
Behavioral & Preference	News media used, alternative information source, news type, reading reason	Nominal	X7–X13
Lifestyle & Access	Lifestyle, access frequency per month, reading duration, pages read, access time, device	Ordinal/Ratio	X14–X19

Source: Compiled by the researcher (2025)

Data Analysis: K-Means Clustering

Data processing was conducted using IBM SPSS Statistics version 27. The analytical procedure comprised four sequential stages: Stage 1 Variable Grouping: Nineteen variables were categorized into demographic and behavioral-preference dimensions for input into the clustering algorithm. Stage 2 Validity and Reliability: Expert Judgement validation with three specialists confirmed content validity. Stage 3 Optimal Cluster Determination: The Elbow Method was applied by computing WCSS for $k = 2, 3, 4,$ and 5 clusters. The 'elbow' point was identified at $k = 3$ based on the most significant WCSS reduction. Stage 4 Cluster Profiling: Final Cluster Centers from SPSS were analyzed to construct detailed behavioral and demographic profiles for each cluster, supplemented by in-depth interview transcription with 12 informants (3 per cluster, plus supervisors).

RESULT AND DISCUSSION

Respondent Demographics Overview

The 100 respondents showed a balanced but slightly female-dominant gender distribution (60% female, 40% male). The majority held bachelor's degrees (S1: 59%) and resided on Java Island (73%). Age distribution reflected a productive workforce profile with modes at 31 years (PNS/civil servants) and 37 years (private employees), alongside a cohort of students aged 19. Detik.com (40%) and Kompas.com (38%) dominated as primary media platforms.

Elbow Method — Determining Optimal Number of Clusters

The Elbow Method analysis revealed the WCSS values across different cluster configurations (Table 3). The most significant WCSS reduction occurred between 2 and 3 clusters (61.08 units), while subsequent reductions were marginal, confirming $k = 3$ as the optimal solution.

Table 2. WCSS Values by Number of Clusters (Elbow Analysis)

Clusters (k)	WCSS Value	Reduction from Previous	Decision
2	1051.721	— Baseline	Baseline
3	990.640	-61.08 (significant)	Optimal (Elbow Point)
4	949.470	-41.17 (diminishing)	Not selected
5	936.975	-12.49 (marginal)	Not selected

Source: SPSS output (2025)

Cluster Membership and Distribution

The K-Means algorithm with $k = 3$ produced the following cluster membership distribution (Table 3), with Cluster 1 representing the largest segment at 63% of respondents.

Table 3. Cluster Membership Distribution

Cluster	Name / Label	N	Percentage
1	Working Women Entertainment Reader	63	63%
2	Executive Millennial National News Interest	26	26%
3	The Lifestyle Trend-Seeker	11	11%
Total		100	100%

Source: SPSS output (2025)

Final Cluster Centers — Key Behavioral Characteristics

Table 4 presents the most discriminating variables from the Final Cluster Centers, revealing the distinct behavioral and demographic profiles of each reader segment.

Table 4. Final Cluster Centers — Selected Key Variables

Variable	Cluster 1: Women Entertainer	Cluster 2: Exec. Millennial	Cluster 3: Lifestyle Seeker
Gender: Female	0.67 (67%)	0.19 (19%)	0.82 (82%)
Age (mode)	31 years (59%)	37 years (35%)	19 years (45%)
Education: S2/Master	0.13 (13%)	0.35 (35%)	0.18 (18%)
Primary Media	Kompas.com (65%)	Detik.com (46%)	CNN.com (36%)
Preferred News Type	Lifestyle (21%), Hiburan (30%)	Nasional (27%), Politik (19%)	Lifestyle (55%)
Content Format	Video (62%)	Teks/Artikel (46%)	Video & Infografis (55%+36%)
Access Frequency	>10x/day (56%)	>20x/day (50%)	>10x/day (45%)
Reading Duration	>5 min (56%)	>10 min (38%)	>2 min (55%)
Access Time	Night 17:01–24:00 (65%)	Day+Night (35%+54%)	Night (73%)
Device	Smartphone (86%)	Smartphone (73%)	Smartphone (91%)
Other Media Used	Instagram (68%)	Instagram+TV (46%+23%)	TikTok (64%)
Reading Motivation	Update & Hiburan (52%)	Info & Opini (35%+19%)	Update/Anti-FOMO (73%)

Source: SPSS output and in-depth interviews (2025)

Cluster Interpretation and Personas

1. Cluster 1 — Working Women Entertainment Reader (63%)

Cluster 1 is dominated by productive-age women (67%) with a characteristic age center of 31 years (59%), the majority married (76%) and educated to bachelor's degree level (S1: 78%). Working predominantly as civil servants (PNS: 27%), this group accesses news primarily at night (65%) via smartphone (86%). Their dominant media is Kompas.com (65%), attracted by its credible yet engaging editorial stance. Key behavioral traits include strong preference for Video format (62%), entertainment and lifestyle content (30% and 21% respectively), and Instagram as the primary alternative media (68%). The primary pain point is intolerance for lengthy text articles — they respond best to aesthetically curated video content.

2. Cluster 2 — Executive Millennial National News Interest (26%)

Cluster 2 represents the most intensive news consumers in the sample. Dominated by professional males (81%) aged approximately 37 years (35%), with the highest educational profile (S2/Master: 35%), and predominantly married (73%) private-sector employees (50%) based on Java (85%), this cluster constitutes the platform's "power users." They access news more than 20 times per day (50%) with reading durations exceeding 10 minutes (38%), consuming 7+ pages per session (38%). Their primary media is Detik.com (46%), favored for breaking news and analytical depth. Focus areas include national news (27%), politics (19%), and economics (15%). The primary pain point is intrusive advertising that disrupts deep information absorption.

3. Cluster 3 — The Lifestyle Trend-Seeker (11%)

Cluster 3 represents the youngest and most visually oriented segment. Strongly dominated by young women (82%) at a characteristic age of 19 years (45%), comprising students (36%), largely unmarried (73%), and residing both on Java and beyond (27% outside Java — the highest geographic spread of all clusters). This group is heavily dependent on TikTok (64%) as their primary information reference. Their content preferences center on lifestyle (55%) and technology (18%), consumed in video format (55%) and infographics (36%). Primary motivation is staying current (anti-FOMO: 73%). Despite lower per-session duration, this cluster drives the 60.9% user growth in trending content channels.

High-Value Segment Identification and Economic Analysis

Based on traffic data analysis (Table 5), Cluster 2 generates the highest economic value through advertising revenue, while Cluster 1 provides economic stability through household purchasing decision power.

Table 5. Traffic Contribution of Top 5 News Media (2024)

Media	Monthly Visits	Primary Cluster	Strategic Value
Detik.com	178.09 million	Cluster 2 (46%)	Highest traffic volume; ad revenue leader
Kompas.com	160.61 million	Cluster 1 (65%)	Stable loyal readership; premium positioning
Tribunnews.com	Variable	Cluster 1 (mixed)	Mass market; mobile-first optimization needed
CNN Indonesia	High visual	Cluster 3 (36%)	Gen Z growth potential; visual-first strategy
TVOneNews.com	Niche	Cluster 2 (partial)	National news authority; live streaming potential

Source: SimilarWeb (2024) and researcher analysis (2025)

Cluster 2's economic primacy is established through three mechanisms: (1) Engagement Value their extreme access intensity (>20x/day) directly multiplies pageviews and advertising impressions; (2) Premium Monetization Potential their high educational profile (35% Master's degree) makes them ideal targets for premium subscription models and high-value advertising; and (3) Market Authority their engagement maintains web traffic rankings globally, preserving advertising rate cards at premium levels.

Cluster 1's economic stability derives from its demographic profile as primary household purchasing decision-makers married professional women with stable income, making them ideal targets for fast-moving consumer goods (FMCG) and lifestyle product advertising with high conversion rates.

Differentiated Online Value Proposition (OVP) Strategies

Based on the pain-value matrix analysis, three differentiated OVP strategies are formulated (Table 6):

Table 6. Differentiated Online Value Proposition per Cluster

Cluster	Pain Point	OVP Strategy	Key Features
Cluster 1 Women Entertainer	Intolerance for long text; time-constrained reading	Efficiency & Clean Experience	News Bites (rapid summaries); clean layout; aesthetic video cards; minimal ad interruption

Cluster	Pain Point	OVP Strategy	Key Features
Cluster 2 Exec. Millennial	Intrusive advertising interrupting deep reading	Premium Depth & Integrity	Distraction-free reading mode; premium subscription; deep analysis articles; data journalism
Cluster 3 Lifestyle Seeker	Stiff text-heavy content; poor visual engagement	Visual & Interactive Trend-Stream	Short-video news (TikTok-style); aesthetic infographics; relatable language; trending topics

Source: Compiled by the researcher based on cluster centers (2025)

Managerial and Theoretical Implications

Managerial Implications: Each of the five media platforms should pursue segment-specific strategies. Kompas.com should prioritize News Bites for Cluster 1 efficiency and introduce premium subscription tiers for Cluster 2. Detik.com should develop Short-Video News for Cluster 3 growth while strengthening analytical depth for Cluster 2 retention. CNN Indonesia should leverage its visual strength and adopt more relatable language to capture Cluster 3. Tribunnews.com and TVOneNews.com should focus on mobile experience optimization and live interactive content for professional readers.

Theoretical Implications: This study makes three significant theoretical contributions: (1) it validates that behavioral segmentation variables access frequency, reading duration, content interaction provide superior audience profiling precision compared to traditional static demographic approaches; (2) it demonstrates the effectiveness of K-Means Clustering as a machine learning tool for identifying 'invisible' market segments in the digital media ecosystem; and (3) it extends the OVP theoretical framework by empirically demonstrating that value propositions must be cluster-differentiated not generic to maximize engagement and monetization in competitive digital media markets.

CONCLUSION

This study successfully applied K-Means Clustering ($k = 3$, Elbow-optimized) to identify three distinct reader segments across Indonesia's top five online news media platforms from a sample of 100 respondents: Cluster 1 Working Women Entertainment Reader (63%) consists of professional women aged ~31, accessing [Kompas.com](https://www.kompas.com) at night via smartphone for entertainment and lifestyle content in video format, and their profile as household purchasing decision-makers makes them economically stable advertising targets; Cluster 2 Executive Millennial National News Interest (26%) comprises male professionals aged ~37, the platform's heaviest users (>20x/day, >10 min duration), consuming national and political news on [Detik.com](https://www.detik.com), and as the segment generating the highest economic value through traffic volume and premium monetization potential, Cluster 2 is identified as the priority strategic target; and Cluster 3 The Lifestyle Trend-Seeker (11%) consists of young women aged ~19 (Gen Z), highly dependent on TikTok, preferring lifestyle content in visual and video formats, and although the smallest segment, they drive disproportionate growth in trending channels (60.9% user increase). Three differentiated OVP strategies are recommended: Efficiency and Clean Experience for Cluster 1 via News Bites features; Premium Depth and Integrity for Cluster 2 via distraction-free reading and subscription models; and Visual and Interactive Trend-Stream for Cluster 3 via TikTok-style short video news. Study limitations include that

the 100-respondent sample is primarily Java-centric (73%), limiting generalizability to outer island populations, and the study does not account for native advertising value or micropayment models; future research should employ longitudinal designs, larger multi-island samples, and comparative clustering algorithms (e.g., Latent Class Analysis) to validate segment stability over time.

REFERENCES

- Asep Syamsul, M. R. (2018). *Jurnalisme kontemporer: Prinsip dan praktik di era digital*. Media Pressindo.
- Garcia-Murillo, M., & MacInnes, I. (2021). Audience segmentation in digital journalism: A case study from Spain. *Digital Journalism*, 9(5), 689–707.
- Ghandour, A. (2019). Crafting a web-unique value proposition using the concept analysis technique. *Global Business and Economics Review*, 21(1), 14–25.
- Goodstats.id. (2024). *10 portal berita online yang paling sering digunakan publik Indonesia 2024*. Goodstats.id.
- Hadiwibowo, A. W. T. (2023). Analisis segmentasi pengunjung mall menggunakan algoritma K-Means clustering. *Jurnal Informatika dan Sains Data*, 4(1), 55–65.
- Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of cluster in K-Means clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1(6), 90–95.
- Kotler, P., & Keller, K. L. (2022). *Manajemen pemasaran* (Edisi ke-14). Erlangga.
- Lestari, R., & Nugroho, A. (2021). Segmentasi mahasiswa berdasarkan perilaku akademik menggunakan algoritma K-Means. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(3), 225–230.
- Putra, D. A., Prasetyo, H., & Rahman, F. (2022). Clustering artikel pada portal berita online menggunakan metode K-Means. *Jurnal Ilmiah Komputer dan Informatika*, 5(2), 101–110.
- Ramkumar, N. (2025). AI-powered social media content optimization: Enhancing engagement, personalization, and efficiency. In *2025 International Conference on Computing Technologies & Data Communication (ICCTDC)* (pp. 1–10).
- Reuters Institute. (2024). *Digital news report 2024*. Reuters Institute for the Study of Journalism.
- Romli, S. M. (2016). *Jurnalistik praktis*. Simbiosia Rekatama Media.
- SimilarWeb. (2024). *Website traffic statistics: Indonesian news media*. SimilarWeb.
- Sugiyono. (2019). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.
- Suparyanto, R. W., & Rosad, R. (2015). *Manajemen pemasaran: Teori dan aplikasi*. Mitra Cendekia.
- Teguh Trianto. (2016). *Pengantar media massa*. Graha Ilmu.
- Vargo, C. J., Hopp, T., & Agarwal, P. (2024). Balancing brand safety and user engagement in a two-sided market: An analysis of content monetization on Reddit. *Journal of Current Issues & Research in Advertising*, 45(2), 242–256.
- We Are Social, & Hootsuite. (2024). *Digital 2024: Indonesia report*. We Are Social.
- Zhao, Y., Zhang, X., & Wang, H. (2022). Data-driven news personalization based on reader behavior: A K-Means clustering approach. *Journal of Media Analytics*, 4(2), 112–129.
- Zhou, L., Fan, M., & Wang, R. (2020). Clustering online readers based on content preference and engagement behavior. *Journal of Data Science*, 18(3), 443–458.