ANALYSIS OF INTERNET USER AGE USING K-MEANS CLUSTERING

Angga Muhamad Gojali[1], Rizki Aji Nugraha[2], Elkin Rilvani[3]
anggagojali369@gmail.com[1] , rizkiajinugraha93@gmail.com[2], elkin.rilvni@pelitabangsa.ac.id[3]
Universitas Pelita Bangsa

ABSTRACT
The digital age has led to a substantial rise in internet usage across all age groups. Understanding user behavior, especially based on age demographics, is vital for optimizing digital service delivery, targeted marketing, and public policy formation. This study utilizes the K-Means clustering algorithm to analyze various age groups and their patterns in internet usage, as documented in prior empirical studies across Indonesia. By aggregating findings from multiple journals, we cluster age-based behavioral tendencies and observe how demographic variables correlate with digital engagement. Our results reveal that internet usage patterns are significantly clustered around specific age brackets, particularly among the 19–25 and 36–45-year cohorts. This clustering reveals differences in motivations, usage frequency, and digital literacy levels. The research supports the effectiveness of K-Means as a reliable model for demographic analysis in sociotechnical studies.
Keywords*: Data Mining, K-Means Clustering, Internet Users, Age Analysis, Digital Behavior.*

INTRODUCTION

In recent decades, the internet has revolutionized communication, education, commerce, and governance. With increased accessibility and affordability, the number of internet users in Indonesia has grown exponentially across all age segments. However, despite broad adoption, usage patterns vary widely depending on demographic factors—particularly age. Older users may prioritize news and communication, while younger demographics often lean toward entertainment, social media, and e-commerce [1].

To make sense of these variations, data mining becomes essential. Among its many techniques, K-Means clustering has emerged as a reliable method for identifying behavioral groupings within datasets. When applied to demographic data, especially age, K-Means can highlight latent patterns and reveal how different age groups interact with digital platforms [2].

This research focuses on mining insights from secondary studies by applying K-Means clustering to summarized survey results from previous journals. The goal is to classify internet users by age and behavior, thereby enabling deeper understanding and better decision-making in digital policy, commercial design, and educational outreach.

RESEARCH METHOD

1. Research Objective

The central goal of this study is to utilize the K-Means clustering algorithm to classify internet user behavior across various age groups using consolidated data from multiple empirical studies. Rather than collecting primary data, this research aggregates published survey data and applies machine learning to discern cluster patterns.

2. Overview of K-Means Clustering

K-Means is an unsupervised machine learning algorithm designed to partition n observations into k clusters, where each observation belongs to the cluster with the nearest mean. The basic steps of the algorithm include:

1. Selecting the number of clusters (k)
2. Randomly initializing k centroids

3. Assigning each data point to the nearest centroid
4. Updating the centroid positions based on current cluster members
5. Repeating the process until convergence is achieved

K-Means is chosen due to its scalability, ease of implementation, and suitability for numeric, demographic datasets such as age.

3. Data Sources

Data used in this study is drawn from 15 Indonesian journals published between 2020 and 2023. Each entry contains:

- Publisher and year
- Research objectives
- Respondent age groups
- Sampling methods
- Behavioral conclusions

These were extracted, standardized, and formatted into tabular form to allow for K-Means clustering application.

4. Clustering Parameters

To determine the optimal number of clusters (k), the Elbow Method was applied using age distribution frequency. The optimal k-value was identified where the within-cluster sum of squares (WCSS) significantly drops before stabilizing.

The age range categories were numerically encoded:

- Group 1: ≤18
- Group 2: 19–25
- Group 3: 26–35
- Group 4: 36–45
- Group 5: >45

RESULTS AND DISCUSSION

1. Cluster Distribution

Based on the Elbow Method and WCSS evaluation, the optimal number of clusters was determined to be k = 3. This segmentation provided the best balance between inter-cluster distinctiveness and intra-cluster similarity.

The three clusters can be interpreted as follows:

- Cluster 1: Early Adopters (Ages ≤25)

These individuals, largely composed of students and young professionals, exhibit high levels of internet usage across entertainment, education, and social media platforms. They tend to use mobile-first access and are comfortable with multi-platform environments.

- Cluster 2: Productive Users (Ages 26–45)

Representing a highly productive digital workforce, this group is distinguished by purpose-driven internet activity—such as digital transactions, e-learning, and online work collaboration. This cluster shows both high frequency and diversity of digital engagement.

- Cluster 3: Late Majority (Ages >45)

Older users display more limited and selective digital behavior, mainly involving messaging applications, health information retrieval, and occasional online shopping. While some show growing confidence in e-commerce and telehealth, digital literacy is more varied.

2. Age Group Behavior Patterns

Tabel 2 . pattern of internet users in each age group:

Age Group	Common Behaviors
≤18	Gaming, YouTube, early exposure to digital tools
19–25	High-frequency use, multitasking, digital learning
26–35	E-commerce, digital productivity tools
36–45	Professional networking, health-seeking behavior
>45	Conservative browsing, health and social apps

3. Analysis of Behavioral Implications

Cluster 1 – Early Adopters

These users are digital natives. They adapt quickly to new platforms and exhibit less concern for privacy or digital fatigue. Their preferences lean toward real-time communication, interactive media, and gamified learning.

Cluster 2 – Productive Users

This group is digitally fluent and goal-oriented. Their high engagement levels are driven by utility—work tasks, finance apps, and digital education. Their digital maturity makes them ideal targets for professional service platforms.

Cluster 3 – Late Majority

Older users, though once reluctant, are gradually integrating internet tools into daily life. The key barrier remains trust and ease of use. Services targeting this segment must emphasize simplicity, safety, and credibility.

4. Cross-study Validation.

The results were compared with primary journal conclusions. Most clustering outcomes aligned with journal-reported dominant age segments. For example, the study from Jurnal Kesehatan Masyarakat clearly identified 46–55 as the most active demographic in health-related browsing, confirming the accuracy of cluster 3's profile [4].

CONCLUSION

This research applied the K-Means clustering algorithm to a secondary dataset aggregated from 15 published Indonesian journals to classify internet users by age. The results reveal three distinct behavioral clusters: digital-native youth, productive middle-aged users, and cautious older adults. Each cluster reflects varying levels of digital literacy, engagement, and purpose in internet usage.

The analysis underscores that internet usage patterns are deeply shaped by age, with implications for policy design, digital platform development, and user training programs. Younger users often favor interactive and entertainment-focused platforms, while older users lean toward utilitarian and communication-based tools. These insights align with previous studies [1][4][6][10], which emphasize the importance of user-centric design and the influence of demographic segmentation in digital behavior.

Educational institutions and businesses should tailor digital solutions to match the behavioral tendencies of each cluster. For example, digital literacy programs aimed at older adults may increase their participation in e-health services, while gamified e-learning tools can further engage younger users and improve retention in digital education environments.

Moreover, this study affirms the strength of unsupervised machine learning techniques, particularly K-Means, in revealing latent patterns in demographic data. The clustering results not only support prior survey-based findings but also present a scalable approach to analyzing user behavior at population levels. While limitations exist in terms of secondary data precision and lack of real-time behavioral variables, the integration of multi-

source academic references enhances the credibility of the analysis.

Future research could incorporate real-time behavioral tracking, explore hybrid models such as K-Means++ or DBSCAN, and analyze additional demographic factors such as gender, education level, and geographic location. Such directions will contribute to a more nuanced and dynamic understanding of internet usage behavior in Indonesia and beyond..

REFERENCES

Additional references from the remaining 9 journals merged based on clustering input data.

Azid, T., et al. "Clustering Analysis on Online Purchasing Behavior Using K-Means." 2021..

Central Bureau of Statistics (BPS). (2023). Proportion of Individuals Owning/Mastering Mobile Phones by Age Group.

Community Pharmacy Journal (Unair). "The Relationship Between Age and Students' Knowledge and Behavior Towards Health Protocols." 2020.

GoodStats. (2023). Latest Data, Profile of Internet Users' Age in Indonesia.

Indonesian Internet Service Providers Association (APJII). (2024). Indonesian Internet Survey Report.

Information Systems Journal (Unikom). "Internet User Prediction Analysis Based on Demographics Using Data Mining." 2021.

Journal of Applied Intelligent Computing (UIN Jakarta). "Application of K-Means for Segmenting Digital Media User Preferences." 2022.

Journal of Early Childhood Education. (2022). Gadget Usage in Early Childhood.

Journal of Statistics and Computing. "Analysis of Digital Trends Based on Age and Region in Indonesia." 2023.

Journal of Technology and Computer Systems (ITS). "Evaluation of Internet Usage Patterns Based on Millennial and Gen Z Generations." 2022.

Katadata. (2023). Majority of Internet Users in Indonesia Are from the Working Age Group.

Pharmaceutics Journal (Unpad). "Analysis of Age and Knowledge of Online Pharmacy Services." 2020.

Potensia Scientific Journal (UNIB). "Digital Media Usage in Early Childhood." 2020.

PREPOTIF Journal of Public Health. "The Relationship Between Age and Public Health Behavior." 2020.

Scientific Journal of Computer Engineering (UPI). "Demographic Segmentation of E-Commerce Users Using Clustering." 2023.

Scientific Journal of Urban and Regional Studies. "Age and Knowledge Regarding Urban Environment." 2020.