JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 629 – 636 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i4.4112
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

CRITERIA ANALYSIS OF COURSE PARTICIPANTS USING K-MEANS: A CASE STUDY OF INET PALEMBANG

Muhammad Rasuandi Akbar[1], Rezania Agramanisti Azdy[2*], Yesi Novaria Kunang[2], Nurul Adha Oktarini Saputri [2]
[1]Informatics Engineering, Bina Darma University
[2] Faculty of Science and Technology, Bina Darma University
email: [*] mrasuandiakbar@gmail.com

Abstract: INET Computer Palembang, as a computer training institution, faces difficulties in understanding participant characteristics due to variations in age, educational background, and chosen course packages. This study aims to analyze participant criteria and group them based on similarities using the K-Means Clustering algorithm. The data used were historical records of course participants from 2022 to 2025. The research process followed the CRISP-DM stages, starting from data cleaning and transformation, determining the optimal number of clusters using the Elbow Method, to evaluating cluster quality with the Davies-Bouldin Index. The implementation was carried out using Python and the scikit-learn library. The results show that the optimal number of clusters is k=5 with a Sum of Squared Errors (SSE) value of 1064.66 and a Davies-Bouldin Index (DBI) score of 0.820, indicating good cluster quality. The resulting clustering provides a structured profile of participants and demonstrates that K-Means is effective in segmenting course participants. These findings are expected to assist the institution in designing more targeted training programs.

Keywords: clustering; data mining; elbow method; k-means; computer course

Abstrak: INET Komputer Palembang sebagai lembaga kursus komputer menghadapi kendala dalam memahami karakteristik peserta karena beragamnya usia, latar belakang pendidikan, dan paket kursus yang diambil. Penelitian ini bertujuan untuk menganalisis kriteria peserta dan mengelompokkan mereka berdasarkan kesamaan atribut menggunakan algoritma K-Means Clustering. Data yang digunakan berupa data historis peserta kursus periode 2022–2025. Proses penelitian mengikuti tahapan CRISP-DM, dimulai dari pembersihan dan transformasi data, penentuan jumlah klaster optimal menggunakan metode Elbow, hingga evaluasi kualitas klaster dengan Davies-Bouldin Index. Implementasi dilakukan menggunakan bahasa pemrograman Python dan pustaka scikit-learn. Hasil penelitian menunjukkan bahwa jumlah klaster optimal adalah k=5 dengan nilai Sum of Squared Errors (SSE) sebesar 1064.66 dan nilai Davies-Bouldin Index (DBI) 0.820, yang mengindikasikan kualitas klaster cukup baik. Klasterisasi yang dihasilkan mampu memberikan gambaran profil peserta secara lebih terstruktur dan membuktikan bahwa metode K-Means efektif dalam segmentasi peserta kursus. Temuan ini diharapkan dapat mendukung pengelola dalam merancang program yang lebih tepat sasaran.

Kata kunci: clustering; data mining; elbow method; k-means; kursus komputer

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 629 – 636 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i4.4112
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

The development of digital technology has made data an important asset in supporting accurate and strategic decision making [1]. Data has a central role because it is able to provide objective information that helps management in evaluating business conditions and formulating solutions [2]. In the context of non-formal educational institutions, course participant data can be utilized to understand their characteristics, interests, and needs in a more measurable way [3]. However, there are still many institutions that do not have a data analysis system that is capable of grouping participants in a structured manner [4].

INET Komputer Palembang, as a computer course institution, faces challenges in understanding the diversity of participants, both in terms of age, educational background, and the type of course package chosen. This variation makes it difficult for management to determine appropriate learning strategies. Without participant segmentation, course program design has the potential to be ineffective because it fails to consider the different needs of each participant group. Maulida et al.[5] expressed a similar thing that the absence of a participant classification system can hinder the development of a curriculum that is appropriate to individual abilities.

One of the relevant approaches to address this problem is the K-Means Clustering method [6]. This method is included in the unsu-pervised learning technique used in data mining to group data based on attribute similarities [7]. Compared to other methods, K-Means is known to be efficient in processing large data sets and is capable of producing clear, easily interpretable clusters. Previous research has also demonstrated the effectiveness of K-Means in education, including for grouping student learning outcomes [8], study program recommendations [9], and mapping of students' extracurricular interests [10].

Based on these conditions, this study was conducted to analyze the criteria for course participants at INET Komputer Palembang using the K-Means algorithm. The goal is to produce clusters that represent participant profiles in a more structured manner, thereby providing useful information for institutions in developing course programs that better suit the needs and characteristics of participants.

.

METHOD

This research was conducted at INET Komputer Palembang, an information technology course and training institution. The research location was chosen because it has a large number of course participants with varying criteria, such as age, education, and course packages, making it relevant for clustering analysis. The research lasted three months, from June 2025 to August 2025, encompassing data collection, processing, modeling, and reporting.

The data used was historical course participant data from 2022–2025, with a total of 763 entries. Data collection was conducted through interviews with INET Komputer administrators, observations of the system, and literature review. The main variables analyzed included age, education level, and course package type.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 629 – 636 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i4.4112
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

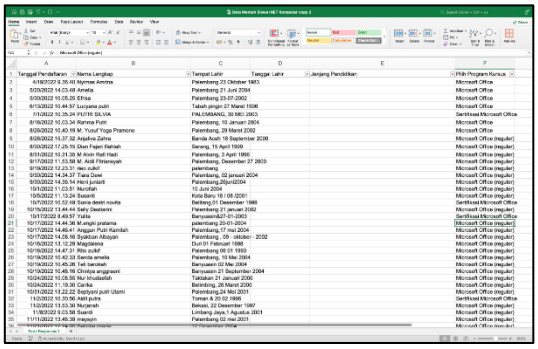

Figure 1. INET Computer Course Participant Data 2022-2025

This research uses a quantitative approach with the CRISP-DM (Cross-Industry Standard Process for Data Mining) method which consists of five stages, namely business understanding, data understanding, data preparation, modeling, and evaluation [11]. An illustration of the CRISP-DM stages can be seen in Figure 1.

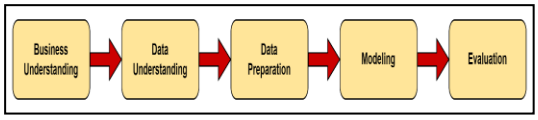

Figure 2. CRISP DM Stages

The data preparation stage involved integration, cleaning of blank and duplicate values, transforming birth dates to ages, converting categorical data to numeric data using label encoding, and standardization with StandardScaler.

The modeling stage was performed using the K-Means Clustering algorithm. The optimal number of clusters was determined using the Elbow method by calculating the Sum of Squared Errors (SSE) for each k value. The SSE formula is shown in equation (1).

$$SSE = \sum_{k=1}^{K} \sum_{x_i \in C_k} ||x - c_k||^2 \quad (1)$$

(1)

Description
SSE = sum of squared errors;
k = number of clusters;
x_i = i-th data point;
c_k = centroid of k-th cluster.

The K-Means algorithm then groups the data based on the distance to the cluster center (centroid) using Euclidean Distance as in equation (2).

$$D_{i,j}\sqrt{(x_{1p} - x_{pq})^2 + (x_{2p} - x_{2q})^2 + \cdots + (x_{rp} - x_{rq})^2} \quad (2)$$

Description:
$D_{i,j}$ = distance between the p-th data point and the q-th cluster center;
x_{rp} = value of the r-th attribute in the p-th data point;
x_{rq} = value of the r-th attribute at the q-th cluster center.

The cluster centers are updated iteratively by calculating the average positions of the cluster members using equation (3).

$$v = \frac{\sum_{p}^{n} = 1\, x_i}{n}; p = 1,2,3, \ldots, n \quad (3)$$

Description:
v = cluster centroid;
x_i = i-th object;
n = number of objects in the cluster.

The evaluation stage is carried out using the Davies-Bouldin Index (DBI), which measures cluster quality based on density within clusters and separation between clusters. A smaller DBI value indicates better clustering results. The DBI formula is shown in equation (4).

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 629 – 636 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i4.4112
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \quad max_{i \neq j}(R_{ij}) \quad (4)$$

Description:
DBI = Davies-Bouldin Index value;
k = number of clusters;
R_{ij} = ratio of the distance between clusters to the average distance within a cluster.

With this procedure, the research can be replicated using similar datasets and the same algorithms, so that the results obtained can be verified and used to understand the characteristics of course participants in a more structured manner.

RESULTS AND DISCUSSION

The results of this study show that the K-Means algorithm is capable of clustering 763 course participants at INET Computer Palembang into five clusters with different characteristics. The analysis was carried out through several stages, starting with descriptive data exploration, determining the optimal number of clusters using the Elbow method, evaluating cluster quality with the Davies-Bouldin Index (DBI), and interpreting the clustering results to reveal the actual needs of the course participants.

To understand the basic profile of course participants before the clustering process is carried out, descriptive analysis is conducted on the course program data, place of birth, and education level. The distribution results are shown in Table 1, Table 2, and Table 3.

Table 1. Course Program Distribution

Course Program	Amount	Percentage
Microsoft Office (Reguler)	401	52,6%
Microsoft Office – Excel	104	13,6%
Microsoft Office (Khusus CPNS)	123	16,1%
AutoCAD	48	6,3%
Pemrograman PHP	14	1,8%
Sertifikasi + Sertifikasi Microsoft Off	21	2,8%
Komputer Fundamental, Web Dev, Grafis	27	3,5%
Others (≤6 participants per program)	25	3,3%
Total	763	100%

Table 2. Distribution of Place of Birth (Top 10)

Place of birth	Amount	Percentage
Palembang	447	58,6%
Banyuasin	38	5,0%
Baturaja	12	1,6%
Jakarta	9	1,2%
Lubuklinggau	9	1,2%
Muara Enim	7	0,9%
unknown	6	0,8%
Tanjung Enim	5	0,7%
Pendopo	5	0,7%

Table 3. Distribution of Educational Levels

Educational level	Amount	Percentage
Collage	297	38,9%
High School/Vocational School/Equivalent	183	24,0%
Other	167	21,9%
Unknown	77	10,1%
Already Working	39	5,1%
Total	763	100%

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 629 – 636 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i4.4112
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Based on Table 1, Microsoft Office courses (regular and CPNS) dominated the participant needs, accounting for over 80% of the total. This confirms that office applications remain the most basic skills needed in the workforce. Table 2 shows that the majority of participants came from Palembang (58.6%), with other regions contributing less. This distribution indicates a predominance of urban participants. Furthermore, Table 3 shows the dominance of university students (38.9%) and high school/vocational school students (24%). This indicates that computer courses are considered important by both students and prospective workers.

Next, the number of clusters. To determine the optimal number of clusters, the Elbow method was used by calculating the Sum of Squared Errors (SSE) for various kkk values. The SSE values are shown in Table 4, while the Elbow graph is shown in Figure 3.

Table 4. SSE Values for Various K Values

k	SSE
2	2412.55
3	1896.42
4	1389.31
5	1064.66
6	1050.21

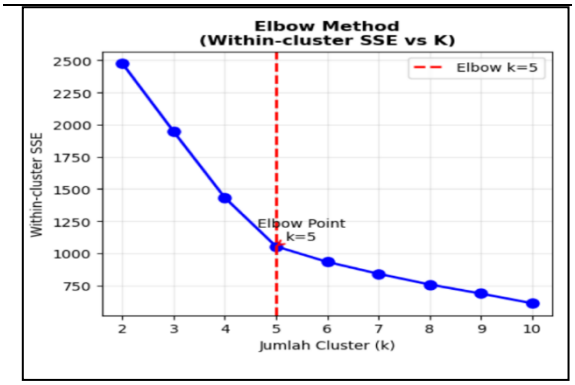

Figure 3. Elbow Method Result Graph

Based on Table 4 and Figure 3, the elbow point occurs at k=5k=5k=5. At this point, the SSE value begins to decline, so the optimal number of clusters is set at five.

Clustering results were validated using the Davies-Bouldin Index (DBI). DBI values for various numbers of clusters are shown in Table 5.

Table 5. DBI Value per Cluster

k	DBI
2	1,294
3	1,107
4	0,932
5	0,820
6	0,845

The best DBI value was obtained at k=5, k=5, k=5, with a score of 0.820. Since the DBI value is <1, the clustering results can be considered quite good, with each cluster having adequate internal cohesiveness and inter-cluster separation. DBI is a reliable evaluation metric for determining cluster quality in non-formal education data.

The distribution of participants across the five clusters using the K-Means method is shown in Table 6.

Table 6. Participant Cluster Distribution (k=5)

Cluster	Proportion	Dominant Characteristics
C0	16,3%	Student from Banyuasin; Microsoft Office package
C1	37,2%	Student from Palembang; Microsoft Office package
C2	0,7%	Workers aged ≥40 years; Excel

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 4, September 2025, hlm. 629 – 636 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i4.4112
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

		focus
C3	15,3%	Palembang students; interested in AutoCAD and preparing for CPNS
C4	30,5%	High school/vocational school graduate; Microsoft Office package

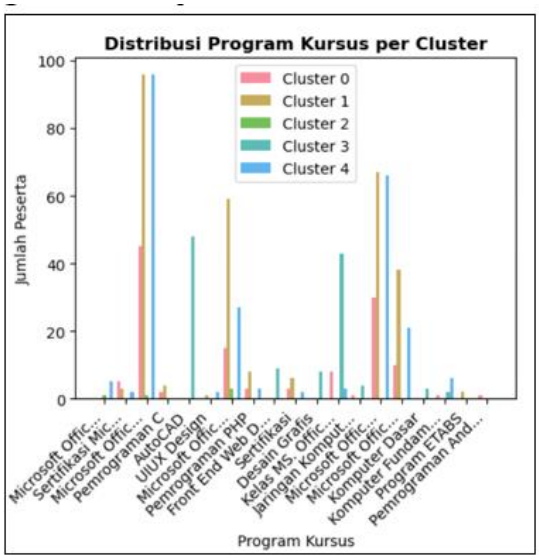

Figure 4. Visualization of course participant clustering results

Table 6 and Figure 3 show that the largest clusters are C1 and C4, both emphasizing Microsoft Office skills. The difference lies in educational background, with C1 dominated by Palembang students, while C4 consists of high school/vocational high school graduates. Cluster C0 represents the student segment from Banyuasin, while C3 represents Palembang students interested in AutoCAD and the Civil Service Candidate (CPNS). Cluster C2, although small, is noteworthy because it contains middle-aged workers with a focus on Excel.

The results of this study provide practical implications for training institutions. First, the significant demand for Microsoft Office underscores the need for tiered training programs, from basic to advanced. Second, participants interested in AutoCAD and CPNS can be directed to more technical and applied programs. Third, the middle-aged worker segment presents an opportunity to design specialized "Excel for Work" classes tailored to industry needs. Academically, this research strengthens the literature which states that data mining-based clustering is effective in mapping the needs of non-formal education participants.

CONCLUSION

This study successfully grouped 763 course participants at INET Komputer Palembang into five clusters using the K-Means algorithm. The clustering results indicated that the primary need for participants was Microsoft Office proficiency, both among university students and high school/vocational school graduates, underscoring the importance of basic computer skills in supporting education and the workforce. Furthermore, other clusters revealed specific segments, such as students who needed AutoCAD and CPNS preparation, and middle-aged workers who focused on Excel as a practical skill.

Scientifically, this study reinforces previous findings that data mining-based clustering methods can provide deeper insights into the needs of non-formal education participants. Practically, the research results can be utilized by course institutions to design more adaptive curricula, such as tiered

Microsoft Office programs, AutoCAD technical classes, and specialized Excel training for workers.

This study is still limited to one course institution in Palembang; therefore, future research could expand the data coverage by involving more institutions or other regions. In addition, testing of other clustering algorithms, such as DBSCAN or Hierarchical Clustering, can also be performed to compare the quality of the results and provide a more comprehensive perspective.

BIBLIOGRAPHY

[1] N. Wijaya, K. Lie, M. Akbar, Q. P. Effendy, and D. F. A. Hariyadi, "Optimalisasi Pemilihan Smartphone Berbasis AI Tahun 2025 Menggunakan Metode Weighted Product dalam Sistem Pendukung Keputusan," *Digital Transformation Technology*, vol. 5, no. 1, pp. 107–114, May 2025, doi: 10.47709/DIGITECH.V5I1.5855.

[2] P. Studi Manajemen and F. Ekonomi dan Bisnis Islam, "PENGARUH KUALITAS DATA TERHADAP INOVASI DALAM MANAJEMEN PEMASARAN," *Musytari : Jurnal Manajemen, Akuntansi, dan Ekonomi*, vol. 6, no. 4, pp. 81–90, Jul. 2024, doi: 10.8734/MUSYTARI.V6I4.4267.

[3] * Dinda, R. Andini, E. Fitrianti, E. A. Lestari, and D. Brutu, "Peran Organisasi Pendidikan di Luar Sekolah dalam Meningkatkan Kualitas Pembelajaran Non-Formal: Studi Kasus di Lembaga Kursus dan Pelatihan," *Jurnal Inovasi, Evaluasi dan Pengembangan Pembelajaran (JIEPP)*, vol. 5, no. 1, pp. 158–163, Apr. 2025, doi: 10.54371/JIEPP.V5I1.794.

[4] A. R. N. Nabella, H. Z. Zahro', and Y. A. Pranoto, "Rancang Bangun Sistem TOEFL Untuk Analisis Kelemahan Peserta Dengan Penerapan Algoritma K-Means Clustering," *Infotek: Jurnal Informatika dan Teknologi*, vol. 8, no. 1, pp. 94–103, Jan. 2025, doi: 10.29408/JIT.V8I1.28260.

[5] V. Maulida, N. Mulyani, and M. F. L. Sibuea, "Sistem Klasifikasi Strata Kelas Peserta Kursus berbasis web menggunakan algoritma K-Means," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, pp. 477–486, Dec. 2024, doi: 10.29408/edumatic.v8i2.27311.

[6] J. Jin, "Student behavior patterns in vocational education big data based on clustering algorithm," *Discover Artificial Intelligence*, vol. 5, no. 1, pp. 1–17, Aug. 2025, doi: 10.1007/S44163-025-00433-3/FIGURES/5.

[7] K. P. Sinaga and M. S. Yang, "A Globally Collaborative Multi-View k-Means Clustering," *Electronics 2025, Vol. 14, Page 2129*, vol. 14, no. 11, p. 2129, May 2025, doi: 10.3390/ELECTRONICS14112129.

[8] W. Kurniawan and R. Kurniawan, "PENERAPAN ALGORITMA K-MEANS CLUSTERING DALAM MENENTUKAN PELUANG MASUK SISWA KE UNIVERSITAS NEGERI," *Jurnal Informatika Teknologi dan Sains (Jinteks)*, vol. 7, no. 1, pp. 386–393, Mar. 2025, doi: 10.51401/JINTEKS.V7I1.5586.

[9] A. Mulyana, Y. Hermawan, and N. J. Saputri, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Rekomendasi Pilihan Program Studi Pada Mahasiswa Baru (Studi Kasus di Institut Bisnis dan Informatika Kesatuan)," *KERNEL: Jurnal Riset Inovasi Bidang Informatika dan Pendidikan Informatika*, vol. 5, no. 1, pp. 60–72, Jul. 2024, doi: 10.31284/J.KERNEL.2024.V5I1.7624.

[10] Muslimah, R. T. Subagio, and V. D. Kartika, "Penerapan Metode K-Means untuk Klasterisasi Minat dan Bakat Siswa terhadap Ekstrakurikuler Sekolah," *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 8, no. 3, pp. 882–897, Aug. 2024, doi: 10.33395/REMIK.V8I3.14008.

[11] G. Gunawan, "DATA MINING USING CRISP-DM PROCESS FRAMEWORK ON OFFICIAL STATISTICS: A CASE STUDY OF EAST JAVA PROVINCE: A case analysis of East Java Province," *Jurnal Ekonomi dan Pembangunan*, vol. 29, no. 2, pp. 183–198, Dec. 2021, doi: 10.14203/JEP.29.2.2021.183-198.