

## Implementasi K-Means Clustering untuk Mengelompokkan Data Produktivitas Penelitian Dosen di Una'im Yapis Wamena

Syarifah<sup>1</sup>, Muhammad Ali<sup>2</sup>, Asman Samad<sup>3</sup>, Fayra Nabillah Azza<sup>4</sup>, Hasriani M<sup>5</sup>, Susiana<sup>6</sup>

<sup>1</sup>Manajemen Retail, Universitas Amal Ilmiah Yapis Wamena, Wamena, Papua Pegunungan

<sup>2</sup>Administrasi Publik, Universitas Amal Ilmiah Yapis Wamena, Wamena, Papua Pegunungan

<sup>3</sup>Magister Manajemen, Universitas Muhammadiyah Makassar, Makassar, Sulawesi Selatan

<sup>4</sup>Ilmu Pemerintahan, Universitas Amal Ilmiah Yapis Wamena, Wamena, Papua Pegunungan

<sup>5</sup>Manajemen Retail, Universitas Amal Ilmiah Yapis Wamena, Wamena, Papua Pegunungan

<sup>6</sup>Manajemen Keuangan Sektor Publik, Universitas Amal Ilmiah Yapis Wamena, Wamena, Papua Pegunungan

Email: <sup>1</sup>syarifah@unaim-wamena.ac.id, <sup>2</sup>muhammadali@unaim-wamena.ac.id, <sup>3</sup>asmansamad748@gmail.com, <sup>4</sup>fayranabillahazza.12@gmail.com, <sup>5</sup>hasrianimuslimin1@gmail.com, <sup>6</sup>susiana@unaim-wamena.ac.id

### Abstrak

Una'im Yapis Wamena sebagai salah satu institusi pendidikan tinggi di Papua memiliki sejumlah dosen dengan beragam aktivitas penelitian. Namun, tantangan yang dihadapi adalah sulitnya mengidentifikasi pola dan karakteristik penelitian dosen secara komprehensif. Tanpa analisis yang tepat, sulit bagi institusi untuk mengetahui profil produktivitas penelitian dosen, mengidentifikasi dosen yang memerlukan pendampingan, atau menentukan fokus pengembangan kapasitas riset. Hal ini dapat berdampak pada kurang optimalnya alokasi sumber daya, dukungan penelitian, dan perencanaan pengembangan penelitian di institusi. Penelitian ini bertujuan untuk membantu mengetahui tingkat produktivitas Dosen dalam penelitian dengan kategori tinggi, sedang, dan rendah. Dengan pengelompokan ini, harapannya Kampus Una'im Yapis Wamena dapat lebih mudah membuat keputusan tentang metode pembinaan yang tepat untuk meningkatkan hasil penelitian Dosen. Metode yang digunakan dalam penelitian ini yaitu K-means Clustering. Data yang dianalisis meliputi 45 Dosen dengan variabel jumlah publikasi, jumlah sitasi, H-index, penelitian yang didanai, dan kolaborasi penelitian. Hasil penelitian ini menunjukkan bahwa tingkat produktivitas penelitian Dosen dapat dikelompokkan menjadi tiga kategori, yaitu: Cluster 1 (Produktivitas Ttinggi) terdiri dari 8 Dosen, Cluster 2 (Produktivitas Sedang) terdiri dari 25 Dosen, dan Cluster 3 (Produktivitas Rendah) terdiri dari 12 Dosen. Proses clustering mencapai konvergensi pada iterasi ke-6 dengan hasil yang stabil. Hasil clustering ini dapat menjadi panduan bagi Una'im Yapis Wamena dalam menentukan kebijakan peningkatan produktivitas penelitian dosen di masa mendatang.

**Kata Kunci:** centroid, data mining, k-mean clustering, penelitian, produktivitas dosen

## Implementation of K-Means Clustering to Group Research Productivity Data of Lecturers at Una'im Yapis Wamena

### Abstract

Una'im Yapis Wamena, as one of the higher education institutions in Papua, has a number of lecturers with various research activities. However, the challenge faced is the difficulty of comprehensively identifying the patterns and characteristics of faculty research. Without proper analysis, it is difficult for the institution to determine the research productivity profile of faculty members, identify those who require mentoring, or determine the focus of research capacity development. This can result in suboptimal allocation of resources, research support, and research development planning within the institution. This study aims to help determine the level of productivity of lecturers in research with high, medium, and low categories. With this grouping, it is hoped that the Una'im Yapis Wamena Campus can more easily make decisions about the appropriate coaching methods to improve the research results of lecturers. The method used in this study is K-means Clustering. The data analyzed included 45 lecturers with variables such as number of publications, number of citations, H-index, funded research, and research collaboration. The results of this study show that the level of faculty research productivity can be grouped into three categories, namely: Cluster 1 (High Productivity) consisting of 8 faculty members, Cluster 2 (Medium Productivity) consisting of 25 faculty members, and Cluster 3 (Low Productivity) consisting of 12 faculty members. The clustering process reached convergence in the 6th iteration with stable results. These clustering results can serve as a guide for Una'im Yapis Wamena in determining policies to increase lecturer research productivity in the future.

**Keywords:** centroid, data mining, K-Means clustering, research, lecturer productivity

## I. PENDAHULUAN

Perkembangan teknologi informasi telah membawa perubahan signifikan dalam pengelolaan data di berbagai institusi, termasuk perguruan tinggi [1]. Institusi pendidikan tinggi menghasilkan volume data yang besar dari berbagai aktivitas akademik, salah satunya adalah data penelitian dosen [2]. Data penelitian dosen mencakup berbagai aspek seperti jumlah publikasi, bidang penelitian, tingkat kolaborasi, sumber pendanaan, dan dampak penelitian [3]. Pengelolaan data penelitian yang efektif menjadi kunci untuk meningkatkan kualitas riset dan mendukung pengambilan keputusan strategis di perguruan tinggi [4].

Una'im Yapis Wamena sebagai salah satu institusi pendidikan tinggi di Papua memiliki sejumlah dosen dengan beragam aktivitas penelitian. Namun, tantangan yang dihadapi adalah sulitnya mengidentifikasi pola dan karakteristik penelitian dosen secara komprehensif. Tanpa analisis yang tepat, sulit bagi institusi untuk mengetahui profil produktivitas penelitian dosen, mengidentifikasi dosen yang memerlukan pendampingan, atau menentukan fokus pengembangan kapasitas riset. Hal ini dapat berdampak pada kurang optimalnya alokasi sumber daya, dukungan penelitian, dan perencanaan pengembangan penelitian di institusi.

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan analitik yang mampu mengekstraksi informasi bermakna dari data penelitian dosen. Data mining merupakan solusi yang tepat karena memiliki kemampuan untuk menggali pola dan pengetahuan dari sekumpulan data yang besar. Data mining adalah proses mengekstraksi informasi berharga dari sejumlah besar data yang sebelumnya tidak diketahui namun dapat dipahami dan bermanfaat untuk mendukung pengambilan keputusan yang strategis. Salah satu teknik data mining yang efektif untuk mengelompokkan data berdasarkan kemiripan karakteristik adalah clustering.

K-Means Clustering merupakan salah satu algoritma clustering yang paling populer dan banyak digunakan dalam berbagai bidang [5]. Metode ini menggunakan pendekatan unsupervised learning yang bertujuan mengelompokkan data dengan karakteristik serupa ke dalam satu cluster yang sama, sementara data dengan karakteristik berbeda dikelompokkan ke dalam cluster lain. K-Means memiliki keunggulan dalam hal kemudahan implementasi, kecepatan komputasi, dan kemampuan mengolah data dalam jumlah besar dengan efisien [6].

Penerapan K-Means Clustering telah terbukti efektif dalam berbagai konteks pendidikan tinggi [7]. Penelitian Nana menggunakan K-Means untuk mengelompokkan mahasiswa berdasarkan prestasi akademik, yang membantu institusi dalam merancang program bimbingan yang tepat sasaran [8]. Dalam konteks penelitian dan karya ilmiah dosen, Mahiswara menerapkan K-Means untuk mengklasifikasikan dosen berdasarkan kinerja publikasi ilmiah, yang memberikan gambaran jelas tentang

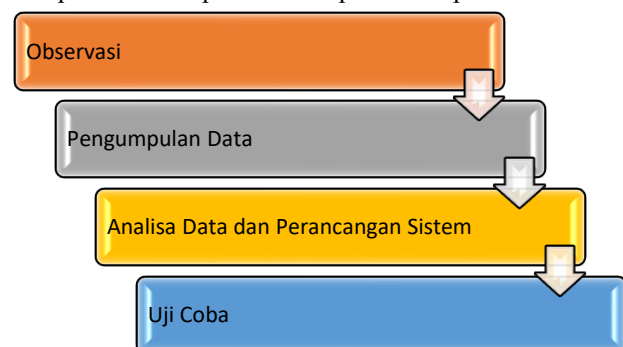
produktivitas riset di institusi tersebut [9]. Selain itu, Asnawi membuktikan efektivitas K-Means dalam pengelompokan data dengan karakteristik beragam [10] [11].

Dalam penelitian ini, penulis menerapkan metode K-Means Clustering untuk mengelompokkan data penelitian dosen di Una'im Yapis Wamena. Data penelitian dosen akan dianalisis dan dikelompokkan berdasarkan berbagai parameter seperti jumlah publikasi, jenis penelitian, dan produktivitas riset. Dengan implementasi metode ini, diharapkan dapat dihasilkan pengelompokan yang memberikan gambaran jelas mengenai profil penelitian dosen, mulai dari kategori sangat produktif, produktif, hingga kurang produktif. Informasi ini akan sangat bermanfaat bagi pimpinan institusi dalam merumuskan kebijakan pengembangan penelitian, memberikan dukungan yang tepat sasaran kepada dosen, dan meningkatkan kualitas serta kuantitas penelitian di Una'im Yapis Wamena

## II. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksploratif. Data penelitian dosen dianalisis menggunakan algoritma K-Means Clustering untuk mengidentifikasi pola dan mengelompokkan dosen berdasarkan produktivitas penelitian. Tahapan penelitian meliputi pengumpulan data, preprocessing data, implementasi K-Means Clustering, evaluasi hasil, dan interpretasi hasil clustering. Objek penelitian adalah data penelitian dosen di Una'im Yapi Wamena yang mencakup periode 5 tahun terakhir (2021-2025). Lokasi penelitian adalah Una'im Yapis Wamena, Papua. Data penelitian diperoleh dari basis data penelitian internal, SINTA, Laporan penelitian dosen, dan database Google Scholar.

Metodologi ini mencakup langkah-langkah sistematis yang diterapkan mulai dari tahap awal hingga akhir penelitian. Setiap tahapan dirancang untuk memastikan bahwa penelitian berjalan secara terstruktur, efisien, dan menghasilkan temuan yang valid serta dapat diandalkan. Melalui metodologi ini, diharapkan penelitian dapat memberikan solusi yang tepat dan aplikatif terhadap permasalahan yang dihadapi, di mana metode ini juga telah banyak diterapkan oleh peneliti sebelumnya [12][13] [14]. Tahapan atau alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Workflow Penelitian

A. Observasi

Melakukan pengamatan langsung terhadap Una'im Yapis Wamena, yang berfungsi sebagai objek penelitian. Melalui observasi ini, peneliti mendapatkan pemahaman yang lebih mendalam mengenai kondisi dan permasalahan yang dihadapi oleh institusi, khususnya terkait produktivitas penelitian dosen.

B. Pengumpulan Data

Proses pengumpulan data dalam penelitian ini mencakup beberapa langkah penting. Pertama, peneliti mengambil data langsung dari sistem informasi penelitian Una'im Yapis Wamena yang relevan dengan topik penelitian. Data ini mencakup informasi publikasi jurnal, sitasi, h-index, penelitian yang didanai, dan kolaborasi penelitian periode 2021-2025. Selain itu, peneliti juga mengumpulkan data referensi dari berbagai sumber literatur seperti buku, jurnal, dan website yang berkaitan dengan metode K-Means Clustering. Data referensi ini digunakan untuk memperkaya landasan teori serta memperkuat analisis dalam penelitian.

C. Analisa Data dan Perancangan Sistem

Pada tahap ini, peneliti melakukan analisis terhadap data yang telah dikumpulkan dan merancang sistem yang sesuai dengan kebutuhan perusahaan. Analisis ini melibatkan pemahaman mendalam terhadap teori-teori yang digunakan dalam penelitian dan penerapan metode yang sesuai. Berikut adalah beberapa poin penting dalam tahap ini:

1. Data Mining

Data mining adalah proses mengekstraksi informasi berharga dari sejumlah besar data yang sebelumnya tidak diketahui namun dapat dipahami dan bermanfaat. Informasi ini kemudian digunakan untuk mendukung keputusan manajemen yang penting. Data mining tidak hanya mengungkap pola-pola tersembunyi dalam data, tetapi juga membantu organisasi dalam memahami tren dan perilaku yang relevan, sehingga memungkinkan pengambilan keputusan yang lebih tepat dan strategis [15].

2. K-Means Clustering

K-Means yaitu salah satu metode clustering yang paling umum digunakan dalam data mining. Metode ini bekerja dengan cara mengelompokkan data ke dalam cluster berdasarkan seberapa mirip karakteristiknya [16]. Clustering yaitu proses membagi satu set objek data menjadi himpunan bagian yang disebut cluster, di mana Satu set objek data dibagi menjadi himpunan bagian yang disebut "cluster" dalam proses yang dikenal sebagai "clustering". Objek-objek dalam satu cluster memiliki kemiripan yang lebih besar satu sama lain daripada objek-objek di cluster lain. Algoritma K-Means bekerja dengan menemukan centroid (titik pusat setiap cluster) dan mengelompokkan data berdasarkan seberapa dekat mereka dengan centroid tersebut [17].

Algoritma K-Means sangat efektif dalam mengelompokkan data karena cepat, mudah digunakan, dan dapat disesuaikan dengan berbagai jenis data. Tujuan utama algoritma ini adalah untuk memaksimalkan kemiripan data dalam satu cluster dan meminimalkan kemiripan antar cluster; fungsi jarak geometris digunakan untuk mengukur kemiripan antara data dan centroid, di mana jarak yang lebih pendek menunjukkan kemiripan yang lebih besar [16]. Dalam penelitian ini, digunakan Euclidean distance dengan rumus sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

d = Distance (Jarak)  
x = Data point pertama  
y = Data point kedua  
xi = Nilai atribut ke-i dari data x  
yi = Nilai atribut ke-i dari data y  
n = Jumlah atribut

Proses klasterisasi dapat dilakukan dengan cara mengidentifikasi data yang akan diklaster, setelah melakukan identifikasi data, selanjutnya dapat menghitung nilai jarak antara data ke-x dan data ke-y. Selanjutnya mencari nilai titik centroid dari klaster ke-k sesuai rumus berikut ini:

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i \quad (2)$$

Keterangan:

$\mu_k$  = Nilai titik centroid cluster ke-k  
 $N_k$  = Banyak data dalam cluster ke-k  
xi = Nilai data ke-i dalam cluster ke-k

Berdasarkan rumus diatas dapat disimpulkan bahwa  $\mu_k$  adalah nilai titik centroid dari cluster ke-K. Dataset yang telah final dapat dilakukan pengolahan data menggunakan Algoritma K-Means. Jumlah nilai K dapat ditentukan dengan membentuk jumlah cluster, lalu mencari titik centroid, dan pengelompokkan data berdasarkan jumlah cluster yang sudah ditentukan. Jika titik centroid berubah maka dilakukan lagi pemilihan titik centroid hingga diketahui tidak ada perubahan pada titik centroid.

D. Uji Coba

Tahap ini, pengujian sistem sedang dilakukan dengan menggunakan metode K-Means untuk memastikan masalah yang ada dapat teratasi dengan baik dan hasil clustering sesuai dengan kebutuhan institusi.

### III. HASIL DAN PEMBAHASAN

A. Analisis Data

Data yang peroleh Una'im Yapis Wamena sebanyak 45 data dosen aktif periode 2021-2025. Tujuan dari analisis ini adalah untuk membagi data tersebut ke dalam tiga cluster yang menggambarkan kategori produktivitas penelitian:

cluster 1 (C1) sebagai produktivitas tinggi, cluster 2 (C2) sebagai produktivitas sedang, dan cluster 3 (C3) sebagai produktivitas rendah. Untuk mencapai tujuan ini, digunakan metode K-Means yang memungkinkan pengelompokan data berdasarkan karakteristik produktivitas penelitian yang berbeda.

Data yang digunakan mencakup 5 variabel sebagai berikut:

1. *Publikasi*: Jumlah publikasi jurnal (nasional dan internasional) periode 2021-2025
2. *Sitasi*: Jumlah total sitasi dari semua publikasi
3. *H-Index*: Indeks produktivitas dan dampak publikasi
4. *Penelitian Didanai*: Jumlah penelitian yang mendapat pendanaan
5. *Kolaborasi*: Jumlah penelitian kolaborasi

Sebelum dilakukan clustering, data dinormalisasi menggunakan Min-Max Normalization agar setiap variabel memiliki skala yang sama (0-1). Rumus normalisasi terdapat pada persamaan 3 berikut.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Keterangan:

$X_{norm}$  = Nilai data setelah normalisasi (rentang 0-1)

X = Nilai yang akan dinormalisasi

$X_{min}$  = Nilai minimum dari variable dalam dataset

$X_{max}$  = Nilai maximum dari variabel dalam dataset

Tabel 1. Statistik Deskriptif Data Penelitian Dosen (Data Asli)

Variabel	Min	Max	Mean	Median	Std Dev
Publikasi	0	18	5,2	4	4,8
Sitasi	0	95	15,6	9	22,3
H-Index	0	8	2,4	2	2,1
Penelitian Didanai	0	7	1,9	1	1,8
Kolaborasi	0	9	2,6	2	2,4

### B. Implementasi K-Means Clustering

Tahap awal dalam penerapan metode K-Means adalah menentukan jumlah cluster yang diinginkan. Dalam penelitian ini, diputuskan untuk membentuk tiga cluster. Langkah selanjutnya adalah menetapkan titik pusat awal (centroid) untuk setiap cluster. Centroid ini berfungsi sebagai titik acuan untuk mengelompokkan data. Pada awalnya, centroid dipilih secara acak dari data yang ada. Pemilihan centroid awal ini penting karena akan mempengaruhi hasil akhir clustering, dan proses iterasi dilakukan untuk memperbarui centroid hingga stabilitas tercapai, yaitu saat tidak ada perubahan signifikan pada posisi centroid.

Centroid awal dipilih dari data dosen dengan karakteristik yang berbeda. Centroid awal (data normalisasi) dapat dilihat pada Tabel 2.

Tabel 2. Centroid Awal Data Normalisasi

Cluster	Publikasi	Sitasi	H-Index	Penelitian Didanai	Kolaborasi
C1	0,889	0,884	0,875	0,857	0,889
C2	0,278	0,289	0,250	0,286	0,222
C3	0,056	0,053	0,125	0,143	0,111

Dalam Tabel 2, C1 mewakili dosen dengan produktivitas tinggi (nilai normalisasi mendekati 1), C2 mewakili produktivitas sedang (nilai normalisasi 0.2-0.3), dan C3 mewakili produktivitas rendah (nilai normalisasi mendekati 0). Selanjutnya menghitung jarak data terhadap titik centroid awal dengan menggunakan Euclidean Distance. Dalam setiap proses perhitungan jarak data ke masing-masing centroid untuk iterasi pertama dilakukan menggunakan data ke-1 sampai ke-5.

Perhitungan Iterasi 1

Perhitungan jarak Dosen-1 terhadap cluster

Jarak ke C1:

$$d1 = \sqrt{[(0.333-0.889)^2 + (0.263-0.884)^2 + (0.375-0.875)^2 + (0.429-0.857)^2 + (0.333-0.889)^2]}$$

$$d1 = \sqrt{0.309 + 0.386 + 0.250 + 0.183 + 0.309}$$

$$d1 = \sqrt{1.437} = 1.199$$

Jarak ke C2:

$$d2 = \sqrt{[(0.333-0.278)^2 + (0.263-0.189)^2 + (0.375-0.250)^2 + (0.429-0.286)^2 + (0.333-0.222)^2]}$$

$$d2 = \sqrt{0.003 + 0.005 + 0.016 + 0.020 + 0.012}$$

$$d2 = \sqrt{0.056} = 0.237$$

Jarak ke C3:

$$d3 = \sqrt{[(0.333-0.056)^2 + (0.263-0.053)^2 + (0.375-0.125)^2 + (0.429-0.143)^2 + (0.333-0.111)^2]}$$

$$d3 = \sqrt{0.077 + 0.044 + 0.063 + 0.082 + 0.049}$$

$$d3 = \sqrt{0.315} = 0.561$$

Dosen-1 masuk ke Cluster 2 (jarak terkecil = 0,237)

Perhitungan jarak Dosen-2 terhadap cluster:

Jarak ke C1:

$$d1 = \sqrt{[(0.444-0.889)^2 + (0.358-0.884)^2 + (0.500-0.875)^2 + (0.571-0.857)^2 + (0.444-0.889)^2]}$$

$$d1 = \sqrt{0.198 + 0.277 + 0.141 + 0.082 + 0.198}$$

$$d1 = \sqrt{0.896} = 0.947$$

Jarak ke C2:

$$d2 = \sqrt{[(0.444-0.278)^2 + (0.358-0.189)^2 + (0.500-0.250)^2 + (0.571-0.286)^2 + (0.444-0.222)^2]}$$

$$d2 = \sqrt{0.028 + 0.029 + 0.063 + 0.081 + 0.049}$$

$$d2 = \sqrt{0.250} = 0.500$$

Jarak ke C3:

$$d3 = \sqrt{[(0.444-0.056)^2 + (0.358-0.053)^2 + (0.500-0.125)^2 + (0.571-0.143)^2 + (0.444-0.111)^2]}$$

$$d3 = \sqrt{0.151 + 0.093 + 0.141 + 0.183 + 0.111}$$

$$d3 = \sqrt{0.679} = 0.824$$

Dosen-2 masuk ke cluster 2 (jarak terkecil 0,500)

Perhitungan jarak Dosen-3 terhadap cluster:

Jarak ke C1:

$$d1 = \sqrt{[(0.111-0.889)^2 + (0.095-0.884)^2 + (0.125-0.875)^2 + (0.143-0.857)^2 + (0.111-0.889)^2]}$$

$$d1 = \sqrt{0.605 + 0.622 + 0.563 + 0.510 + 0.605}$$

$$d1 = \sqrt{2.905} = 1.704$$

Jarak ke C2:

$$d2 = \sqrt{[(0.111-0.278)^2 + (0.095-0.189)^2 + (0.125-0.250)^2 + (0.143-0.286)^2 + (0.111-0.222)^2]}$$

$$d2 = \sqrt{0.028 + 0.009 + 0.016 + 0.020 + 0.012}$$

$$d2 = \sqrt{0.085} = 0.292$$

Jarak ke C3:

$$d3 = \sqrt{[(0.111-0.056)^2 + (0.095-0.053)^2 + (0.125-0.125)^2 + (0.143-0.143)^2 + (0.111-0.111)^2]}$$

$$d3 = \sqrt{0.003 + 0.002 + 0 + 0 + 0}$$

$$d3 = \sqrt{0.005} = 0.071$$

Dosen -3 masuk ke cluster 3 (jarak terkecil = 0,071)

Perhitungan jarak Dosen-4 terhadap cluster:

Jarak ke C1:

$$d1 = \sqrt{[(0.833-0.889)^2 + (0.800-0.884)^2 + (0.750-0.875)^2 + (0.714-0.857)^2 + (0.778-0.889)^2]}$$

$$d1 = \sqrt{0.003 + 0.007 + 0.016 + 0.020 + 0.012}$$

$$d1 = \sqrt{0.058} = 0.241$$

Jarak ke C2:

$$d2 = \sqrt{[(0.833-0.278)^2 + (0.800-0.189)^2 + (0.750-0.250)^2 + (0.714-0.286)^2 + (0.778-0.222)^2]}$$

$$d2 = \sqrt{0.308 + 0.373 + 0.250 + 0.183 + 0.309}$$

$$d2 = \sqrt{1.423} = 1.193$$

Jarak ke C3:

$$d3 = \sqrt{[(0.833-0.056)^2 + (0.800-0.053)^2 + (0.750-0.125)^2 + (0.714-0.143)^2 + (0.778-0.111)^2]}$$

$$d3 = \sqrt{0.604 + 0.558 + 0.391 + 0.326 + 0.445}$$

$$d3 = \sqrt{2.324} = 1.525$$

Dosen-4 masuk ke cluster 1 (jarak terkecil = 0,241)

Perhitungan jarak Dosen-5 terhadap cluster:

Jarak ke C1:

$$d1 = \sqrt{[(0.167-0.889)^2 + (0.147-0.884)^2 + (0.250-0.875)^2 + (0.286-0.857)^2 + (0.222-0.889)^2]}$$

$$d1 = \sqrt{0.521 + 0.543 + 0.391 + 0.326 + 0.445}$$

$$d1 = \sqrt{2.226} = 1.492$$

Jarak ke C2:

$$d2 = \sqrt{[(0.167-0.278)^2 + (0.147-0.189)^2 + (0.250-0.250)^2 + (0.286-0.286)^2 + (0.222-0.222)^2]}$$

$$d2 = \sqrt{0.012 + 0.002 + 0 + 0 + 0}$$

$$d2 = \sqrt{0.014} = 0.118$$

Jarak ke C3:

$$d3 = \sqrt{[(0.167-0.056)^2 + (0.147-0.053)^2 + (0.250-0.125)^2 + (0.286-0.143)^2 + (0.222-0.111)^2]}$$

$$d3 = \sqrt{0.012 + 0.009 + 0.016 + 0.020 + 0.012}$$

$$d3 = \sqrt{0.069} = 0.263$$

Dosen-5 masuk ke cluster 2 (jarak terkecil = 0,118)

Berdasarkan hasil perhitungan pada tahap iterasi sebelumnya, selanjutnya dilakukan proses pengelompokan data. Pada tahap ini, data ke-1 hingga data ke-45 dikelompokkan ke dalam kluster C1, C2, atau C3 berdasarkan nilai jarak terkecil antara masing-masing data dengan pusat kluster (centroid) C1, C2, dan C3. Data akan menjadi anggota kluster yang memiliki jarak paling minimal terhadap centroid tersebut.

#### Menghitung Centroid Baru Iterasi 2

Setelah proses pengelompokan pada iterasi pertama selesai, langkah selanjutnya adalah menghitung centroid baru untuk iterasi ke-2. Centroid baru diperoleh dengan cara menghitung nilai rata-rata (mean) dari seluruh data yang menjadi anggota pada masing-masing cluster. Berdasarkan hasil pengelompokan, jumlah anggota pada setiap cluster adalah sebagai berikut:

- Cluster 1 memiliki anggota: 8 dosen
- Cluster 2 memiliki anggota: 24 dosen
- Cluster 3 memiliki anggota: 13 dosen

Perhitungan centroid baru pada C1 Variabel Publikasi:

Rata-rata Publikasi C1

$$= (0.833 + 0.889 + 0.778 + 0.944 + 0.722 + 0.833 + 0.889 + 0.806) / 8$$

$$= 6.694 / 8$$

$$= 0.837$$

Tabel 3. Centroid Baru Iterasi 2

Cluster	Publikasi	Sitasi	H-Index	Penelitian Didanai	Kolaborasi
C1	0,837	0,821	0,797	0,768	0,819
C2	0,289	0,235	0,281	0,304	0,267
C3	0,078	0,065	0,115	0,126	0,094

Selanjutnya, proses iterasi dilakukan secara berulang hingga memenuhi kondisi perhentian, yaitu ketika perubahan antara centroid sebelumnya dan centroid baru sudah tidak signifikan, dengan nilai ambang batas (threshold) kurang dari 0,001. Pada penelitian ini, proses iterasi berhenti pada iterasi ke-6, karena centroid yang dihasilkan telah konvergen dan tidak mengalami perubahan yang berarti.

Tabel 4. Centroid Akhir Iterasi 6

Cluster	Publikasi	Sitasi	H-Index	Penelitian Didanai	Kolaborasi
C1	0,845	0,832	0,813	0,779	0,825
C2	0,298	0,247	0,294	0,312	0,276
C3	0,072	0,061	0,108	0,119	0,089

#### C. Hasil Pengelompokan

Berdasarkan hasil analisa data, maka data penelitian dosen dapat dikelompokkan pada masing-masing cluster.

Tabel 5. Hasil Akhir Pengelompokan Data Penelitian Dosen

Cluster	Publikasi	Sitasi	H-Index	Penelitian Didanai	Kolaborasi
C1	0,845	0,832	0,813	0,779	0,825
C2	0,298	0,247	0,294	0,312	0,276
C3	0,072	0,061	0,108	0,119	0,089

Tabel 6. Distribusi Hasil Clustering

No	Kode Dosen	Jarak Terdekat	Cluster
1	DSN-001	0.237	Cluster 2
2	DSN-002	0.423	Cluster 2
3	DSN-003	0.071	Cluster 3
4	DSN-004	0.241	Cluster 1
5	DSN-005	0.118	Cluster 2
6	DSN-006	0.089	Cluster 3
7	DSN-007	0.356	Cluster 2
8	DSN-008	0.198	Cluster 1
9	DSN-009	0.284	Cluster 2
10	DSN-010	0.145	Cluster 2
11	DSN-011	0.412	Cluster 2
12	DSN-012	0.267	Cluster 3
13	DSN-013	0.094	Cluster 1
14	DSN-014	0.176	Cluster 2
15	DSN-015	0.329	Cluster 2
16	DSN-016	0.387	Cluster 2
17	DSN-017	0.156	Cluster 1
18	DSN-018	0.219	Cluster 2
19	DSN-019	0.298	Cluster 2
20	DSN-020	0.445	Cluster 3
21	DSN-021	0.103	Cluster 2
22	DSN-022	0.372	Cluster 1
23	DSN-023	0.189	Cluster 3
24	DSN-024	0.127	Cluster 2
25	DSN-025	0.334	Cluster 2
26	DSN-026	0.401	Cluster 2
27	DSN-027	0.245	Cluster 2
28	DSN-028	0.167	Cluster 1
29	DSN-029	0.428	Cluster 2
30	DSN-030	0.312	Cluster 2
31	DSN-031	0.082	Cluster 3
32	DSN-032	0.203	Cluster 1
33	DSN-033	0.365	Cluster 2
34	DSN-034	0.278	Cluster 2
35	DSN-035	0.134	Cluster 3
36	DSN-036	0.289	Cluster 2
37	DSN-037	0.419	Cluster 2
38	DSN-038	0.178	Cluster 1
39	DSN-039	0.098	Cluster 3
40	DSN-040	0.356	Cluster 2
41	DSN-041	0.223	Cluster 2
42	DSN-042	0.115	Cluster 3
43	DSN-043	0.215	Cluster 1
44	DSN-044	0.392	Cluster 2
45	DSN-045	0.112	Cluster 3

Dari hasil perhitungan di atas, maka dapat dijabarkan:

- 1) Cluter 1 (Produktivitas Tinggi) sebanyak 8 Dosen 17,8% dengan rata-rata sebagai berikut: 15-18

publikasi, 70-95 sitasi, h-index 6-8, penelitian didanai 5-7, kolaborasi 7-9.

- 2) Cluster 2 (Produktivitas Sedang) sebanyak 25 Dosen 55,6% dengan rata-rata sebagai berikut: 4-7 publikasi, 10-30 sitasi, h-index 2-4, penelitian didanai 1-3, kolaborasi 2-4.
- 3) Cluster 3 (Produktivitas Rendah) sebanyak 12 Dosen 26,6% dengan rata-rata sebagai berikut: 0-2 publikasi, 0-5 sitasi, h-index 0-1, penelitian didanai 0,1, kolaborasi 0-1.

#### D. Karakteristik Setiap Cluster

*Cluster 1 Produktivitas Tinggi:* Kategori ini meliputi dosen yang sangat produktif dalam melakukan penelitian dan publikasi. Dosen pada kategori ini secara konsisten menghasilkan publikasi berkualitas, memiliki sitasi yang tinggi, dan aktif dalam kolaborasi penelitian baik internal maupun eksternal. Institusi harus mempertahankan produktivitas dosen ini dengan memberikan dukungan berkelanjutan seperti akses ke konferensi internasional, dana penelitian yang memadai, dan pengurangan beban mengajar.

*Cluster 2 Produktivitas Sedang:* Kategori ini mencakup dosen yang memiliki produktivitas cukup baik tetapi masih dapat ditingkatkan. Dosen pada kategori ini menunjukkan potensi untuk berkembang ke kategori produktivitas tinggi dengan pembinaan yang tepat. Strategi pembinaan yang dapat diterapkan antara lain: workshop penulisan artikel ilmiah, mentoring dari dosen produktif, dukungan dana penelitian, dan fasilitasi kolaborasi penelitian.

*Cluster 3 Produktivitas Rendah:* Kategori ini adalah dosen dengan produktivitas penelitian yang rendah atau bahkan tidak ada publikasi sama sekali. Dosen pada kategori ini memerlukan perhatian khusus dan program pembinaan intensif. Beberapa faktor yang mungkin mempengaruhi antara lain: kurangnya pemahaman tentang metodologi penelitian, keterbatasan waktu karena beban mengajar yang tinggi, kurangnya motivasi, atau kesulitan dalam menulis artikel ilmiah. Institusi perlu merancang program khusus seperti: pelatihan metodologi penelitian, pendampingan menulis artikel, pemberian insentif, dan pengurangan beban administratif sementara.

#### E. Validasi Hasil Clustering

Untuk mengukur kualitas hasil clustering, dilakukan evaluasi menggunakan beberapa metrik:

1. *Silhouette Score:* 0.487 (skala -1 sampai 1, semakin mendekati 1 semakin baik). Nilai ini menunjukkan bahwa kualitas clustering cukup baik dan data terkelompok dengan baik.
2. *Davies-Bouldin Index:* 0.856 (semakin kecil semakin baik). Nilai ini menunjukkan pemisahan antar cluster yang cukup baik.
3. *Inertia (Within-Cluster Sum of Squares):* 52.34. Nilai ini menunjukkan kompaknya data dalam setiap cluster.

Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh Dutt et al. (2017) yang menggunakan K-Means untuk menganalisis performa akademik mahasiswa

[12]. Dalam penelitian tersebut, K-Means terbukti efektif dalam mengidentifikasi kelompok mahasiswa yang memerlukan intervensi akademik. Begitu pula penelitian ini berhasil mengidentifikasi kelompok dosen yang memerlukan pembinaan khusus.

Penelitian oleh Ramadhani & Kurniabudi (2019) juga menunjukkan bahwa K-Means dapat digunakan secara efektif untuk clustering data akademik dengan hasil yang memuaskan [13]. Dalam penelitian mereka, K-Means berhasil mengelompokkan mahasiswa berdasarkan prestasi akademik dengan akurasi yang tinggi, sama halnya dengan penelitian ini yang berhasil mengelompokkan dosen berdasarkan produktivitas penelitian.

Muningsih et al. (2021) dalam penelitiannya tentang clustering provinsi menggunakan K-Means dengan optimasi Davies-Bouldin Index menunjukkan bahwa pemilihan jumlah cluster yang tepat sangat penting untuk hasil yang optimal [8]. Penelitian ini menggunakan pendekatan serupa dengan menentukan  $K=3$  berdasarkan karakteristik data dan kebutuhan institusi.

Berbeda dengan penelitian sebelumnya yang lebih fokus pada mahasiswa atau data geografis, penelitian ini memberikan kontribusi baru dalam penerapan K-Means untuk analisis produktivitas penelitian dosen, khususnya di wilayah Papua yang memiliki karakteristik unik dalam hal keterbatasan sumber daya dan akses informasi. Berdasarkan hasil clustering, Una'im Yapis Wamena dapat merancang strategi yang berbeda untuk setiap cluster:

#### *Strategi untuk Cluster 1 (Produktivitas Tinggi):*

1. Memberikan reward dan recognition atas prestasi penelitian
2. Menjadikan role model dan mentor bagi dosen lain
3. Memfasilitasi untuk mengikuti konferensi internasional
4. Memberikan alokasi dana penelitian yang lebih besar
5. Mengurangi beban mengajar untuk fokus pada penelitian
6. Mendorong untuk mempublikasikan di jurnal internasional bereputasi

#### *Strategi untuk Cluster 2 (Produktivitas Sedang):*

1. Program mentoring dengan dosen dari Cluster 1
2. Workshop penulisan artikel ilmiah berkala
3. Fasilitasi pembentukan kelompok penelitian
4. Memberikan seed grant untuk penelitian
5. Pendampingan dalam pengajuan proposal hibah
6. Pelatihan metodologi penelitian lanjutan

#### *Strategi untuk Cluster 3 (Produktivitas Rendah):*

1. Analisis mendalam faktor penghambat produktivitas
2. Program pembinaan intensif dan personal coaching
3. Pelatihan dasar metodologi penelitian
4. Pengurangan beban mengajar dan administratif
5. Pairing dengan dosen senior untuk kolaborasi
6. Memberikan target bertahap yang realistis
7. Penyediaan akses ke literatur dan software penelitian

## IV. KESIMPULAN

Metode K-Means Clustering terbukti secara efektif dapat digunakan untuk analisis produktivitas penelitian dosen di Una'im Yapis Wamena. Dengan menggunakan metode ini, dosen dapat dikelompokkan ke dalam tiga kategori berdasarkan tingkat produktivitas penelitian: *Produktivitas Tinggi*, *Produktivitas Sedang*, dan *Produktivitas Rendah*. Dalam penelitian ini, proses iterasi K-Means mencapai kondisi optimal pada iterasi ke-6, di mana tidak ada lagi perubahan signifikan pada centroid yang menunjukkan bahwa hasil clustering telah stabil. Hasil akhirnya menunjukkan bahwa:

1. 8 dosen (17.8%) termasuk dalam kategori Produktivitas Tinggi dengan rata-rata 15-18 publikasi
2. 25 dosen (55.6%) termasuk dalam kategori Produktivitas Sedang dengan rata-rata 4-7 publikasi
3. 12 dosen (26.6%) dikategorikan sebagai Produktivitas Rendah dengan rata-rata 0-2 publikasi

Evaluasi kualitas clustering menggunakan Silhouette Score menghasilkan nilai 0.487 dan Davies-Bouldin Index 0.856, yang menunjukkan bahwa pengelompokan data cukup baik dan cluster terpisah dengan jelas. Dengan adanya hasil clustering ini, Una'im Yapis Wamena dapat menggunakan informasi tersebut untuk mengoptimalkan strategi peningkatan produktivitas penelitian dosen. Institusi dapat merancang program pembinaan yang berbeda untuk setiap cluster: memberikan reward untuk dosen produktif, program mentoring untuk dosen produktivitas sedang, dan pembinaan intensif untuk dosen produktivitas rendah. Hasil penelitian ini sejalan dengan penelitian sebelumnya yang menunjukkan efektivitas K-Means dalam clustering data akademik. Dengan cara ini, institusi dapat memaksimalkan produktivitas penelitian dengan strategi yang tepat sasaran, sekaligus meningkatkan kualitas dan reputasi institusi. Penelitian ini memberikan kontribusi penting dalam penerapan data mining untuk manajemen pendidikan tinggi, khususnya di wilayah Papua yang memiliki tantangan unik dalam pengembangan kapasitas penelitian dosen.

## REFERENSI

- [1] A. A. Zulfa and O. Arifudin, "Peran Sistem Informasi Akademik Berbasis Web Dalam Upaya Meningkatkan Efektivitas Dan Efisiensi Pengelolaan Akademik Di Perguruan Tinggi," 2025.
- [2] P. Hendikawati, N. A. S. Harwanti, W. Wardono, A. Prabowo, M. D. Zahra, and W. R. Saefurrochman, "Pembelajaran Big Data Di Perguruan Tinggi: Potensi Masa Depan, Faktor Pendukung Dan Penghambat," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 3, pp. 2232–2241, Aug. 2025, doi: 10.29100/jupi.v10i3.7711.
- [3] M. Abrar, "Inovasi Dalam Riset Akademik: Strategi Meningkatkan Kualittas dan Relevansi Penelitian Perguruan Tinggi," *ACADEMIA: Jurnal Inovasi Riset Akademik*, vol. 5, no. 2, pp. 51–66, May 2025.

- 
- [4] Syarifah, F. Nabillah Azza, and I. Misrah, "Optimalisasi Kinerja Perguruan Tinggi Melalui Strategi Manajemen Publikasi Ilmiah Berbasis Science and Technology Index SINTA," *Jurnal Mengabdikan dari Hati*, vol. 4, no. 2, pp. 47–56, Aug. 2025.
- [5] F. Khalish, N. M. Piranti, and O. Martadireja, "Implementasi Data Mining Menggunakan Teknik Clustering dengan Metode K-Means," *JlIP: Jurnal Ilmiah Ilmu Pendidikan*, vol. 8, no. 5, pp. 5392–5397, May 2025, [Online]. Available: <http://Jiip.stkipyapisdompnu.ac.id>
- [6] A. Pratama Putra, J. Tshivana, and E. Rilvani, "Perbandingan Teoritis dan Eksperimen Algoritma K-Means dan K-Medoids Dalam Klasterisasi Data," *Kohesi: Jurnal Multidisiplin Saintek*, vol. 10, no. 2, pp. 1–24, 2025, doi: 10.8734/Kohesi.v1i2.365.
- [7] S. Trisya Amanda, K. Berlianindita, A. Jahid Alfarizi, and W. A. Arifin, "Analisis Komparatif SVM dan K-Means dalam Data Mining Untuk Promosi Perguruan Tinggi," *Jurnal Rekayasa Informasi*, vol. 5, no. 2, pp. 87–94, Jul. 2025.
- [8] N. Suarna, N. Rahaningsih, and A. A. Suarna, "Optimalisasi Prestasi Akademik Siswa Melalui Pengelompokan Indeks Prestasi dengan K-Means Clustering," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 4, no. 2, pp. 198–207, May 2025, doi: 10.69916/jkbt.v4i2.321.
- [9] K. Mahiswara, F. Refariansyah, and D. Oktaviani, "Pemetaan Keuangan dan Produktivitas Perguruan Tinggi di Indonesia Berbasis Publikasi Artikel Ilmiah," *Majalah Ilmiah Unikom*, vol. 17, no. 1, pp. 35–42, Sep. 2019.
- [10] Y. P. Putra and R. Nuari, "Application of K-Means Algorithm to Cluster Student's Reading Patterns in the Digital Age," *Jurnal Inovtek Polbeng - Seri Informatika*, vol. 10, no. 1, pp. 320–331, Mar. 2025.
- [11] M. F. Asnawi, N. Fitriyanto, and M. A. Pamoengkas, "Sistem Pendukung Keputusan Kinerja Dosen Menggunakan Sistem Kecerdasan Buatan Berbasis Algoritma K-Means Clustering," *TECHNOMEDIA: Informatics and Computer Science*, vol. 1, no. 2, pp. 67–73, Jul. 2024, doi: 10.58641.
- [12] A. Rahmawati, M. Fadlan, A. Anto, "Implementasi K-Means Clustering Untuk Mengelompokkan Data Sparepart Alat Berat," *Journal of Big Data Analytic and Artificial Intelligence*, vol. 7, no. 2, pp. 33–38.
- [13] N. Cahaya Kamila, M. Hafid, "Pemanfaatan Data Mining untuk Mengelompokkan Data Penjualan Menggunakan Algoritma K-Medoids," *Journal of Big Data Analytic and Artificial Intelligence*, vol. 7, no. 1, pp. 8–13, 2024.
- [14] T. Martuti, E. Tjatur Putri, R. Gusmana, "Penerapan Clustering pada Sistem Pembagian Kelas Siswa Menggunakan Algoritma K-Means," *Journal of Big Data Analytic and Artificial Intelligence*, vol. 6, no. 2, pp. 17–24, 2025.
- [15] L. A. Supriyono *et al.*, *Buku Ajar Big Data dan Data Mining: Konsep, Metodologi, dan Aplikasi*. PT. Sonpedia Publishing Indonesia, 2025.
- [16] K. Handoko, S. Astiti, W. Novrian, T. N. Putra, and D. Darmansah, *K-Means Clustering Menggunakan Aplikasi Rapidminer*. PT. Sonpedia Publishing Indonesia, 2025.
- [17] A. Muis *et al.*, *DATA MINING: Konsep, Metode, dan Aplikasi*. Faaslib Serambi Media, 2025.
-