

IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN JENIS PROYEK VIDEO BERDASARKAN KARAKTERISTIK DI AGENCY PT LOGEEKA GRUP ID

Muhammad Surya Candra¹⁾, Thomas Afrizal²⁾, Umar Wirantasa³⁾

Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Indraprasta PGRI, Jakarta^{1) 2) 3)}

suryacandra820@gmail.com¹⁾, thomztaurus.it@gmail.com²⁾, wirantasaumar@gmail.com³⁾

Abstrak

Agency produksi video menghadapi tantangan krusial dalam pengelolaan proyek yang bervariasi dari segi kompleksitas, anggaran, dan durasi pengerjaan. PT Logeeka Grup ID sebagai salah satu pelaku industri kreatif di Jakarta mengalami kendala dalam distribusi sumber daya tim produksi akibat tidak adanya sistem klasifikasi proyek yang terstandarisasi. Penelitian ini merancang dan membangun aplikasi web berbasis algoritma K-Means Clustering untuk mengelompokkan 30 data proyek video tahun 2024 ke dalam tiga kluster berdasarkan nilai budget (juta rupiah) dan durasi produksi (hari). Data dinormalisasi menggunakan metode Min-Max Normalization sebelum proses iterasi K-Means dijalankan. Sistem dibangun menggunakan PHP dan MySQL, dilengkapi modul kelola proyek, kelola tim, kelola budget, proses clustering otomatis, serta laporan hasil pengelompokan. Hasil iterasi ke-2 menghasilkan kluster yang stabil: Kluster 1 (Proyek Premium) mencakup proyek berbudget tinggi dan durasi panjang, Kluster 2 (Proyek Standard) untuk proyek menengah, dan Kluster 3 (Proyek Quick) untuk proyek berskala kecil. Evaluasi menggunakan Silhouette Score dan Davies-Bouldin Index mengkonfirmasi kualitas pengelompokan yang baik. Sistem terbukti mampu meningkatkan efisiensi perencanaan alokasi sumber daya dan estimasi anggaran di lingkungan agency kreatif.

Kata kunci: K-Means Clustering, Produksi Video, Agency Kreatif, Data Mining, Min-Max Normalization

1. Pendahuluan

Industri kreatif digital di Indonesia tumbuh pesat seiring meningkatnya kebutuhan konten visual dari berbagai sektor bisnis. Berdasarkan data Badan Ekonomi Kreatif (Bekraf), subsektor film, animasi, dan video menyumbang lebih dari 7% terhadap total nilai Produk Domestik Bruto (PDB) ekonomi kreatif nasional. Agency produksi video seperti PT Logeeka Grup ID turut merasakan lonjakan permintaan proyek yang mencakup spektrum sangat luas, mulai dari TV Commercial berskala besar hingga konten media sosial yang diselesaikan dalam hitungan hari.

Beragamnya jenis proyek ini menimbulkan tantangan operasional yang signifikan. Tanpa sistem kategorisasi yang sistematis, manajer proyek harus mengandalkan intuisi dan pengalaman pribadi untuk menentukan alokasi tim, memperkirakan anggaran, serta menyusun jadwal produksi. Penelitian Suryadi dan Mulyana (2021) mencatat bahwa 68% agency kreatif di Indonesia mengalami kendala serius terkait

estimasi waktu dan biaya yang tidak akurat, yang berdampak langsung pada pembengkakan biaya produksi hingga 30-40% dari nilai kontrak awal dan keterlambatan penyerahan yang merusak kepercayaan klien [1].

Data mining menawarkan pendekatan berbasis data untuk mengatasi permasalahan tersebut. Algoritma K-Means Clustering sebagai salah satu metode unsupervised learning terbukti efektif mengelompokkan objek-objek data berdasarkan kemiripan karakteristiknya tanpa memerlukan label kategori sebelumnya [2]. Penerapan K-Means pada portofolio proyek historis memungkinkan sistem untuk secara otomatis menemukan pola tersembunyi dan mengelompokkan proyek ke dalam kluster-kluster yang bermakna berdasarkan variabel numerik seperti budget dan durasi produksi.

Beberapa studi terdahulu telah mengeksplorasi penerapan K-Means dalam konteks manajemen proyek dan segmentasi bisnis. Kusumawati & Wibowo (2022)

menerapkan K-Means untuk segmentasi pelanggan pada platform e-commerce dan memperoleh akurasi pengelompokan yang memuaskan [3]. Andini & Prasetya (2021) menunjukkan bahwa kombinasi Elbow Method dan Silhouette Score merupakan pendekatan yang andal untuk menentukan jumlah kluster optimal pada data bisnis [4]. Namun, aplikasi spesifik K-Means untuk pengelolaan proyek di industri produksi video masih sangat terbatas dalam literatur akademis Indonesia.

Berdasarkan identifikasi masalah tersebut, penelitian ini bertujuan: (1) merancang dan membangun sistem web berbasis PHP dan MySQL yang mengimplementasikan algoritma K-Means untuk pengelompokan proyek video; (2) menerapkan proses normalisasi Min-Max dan iterasi K-Means pada data proyek nyata PT Logeeka Grup ID tahun 2024; serta (3) mengevaluasi kualitas hasil clustering menggunakan metrik Silhouette Score dan Davies-Bouldin Index. Sistem yang dihasilkan diharapkan menjadi alat bantu pengambilan keputusan operasional yang meningkatkan efisiensi dan akurasi perencanaan proyek secara terukur.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode eksperimental komputasional. Objek penelitian adalah data proyek video PT Logeeka Grup ID periode Januari hingga Desember 2024, berjumlah 30 entri proyek yang terdokumentasi dalam sistem pencatatan internal perusahaan. Tahapan penelitian mengikuti kerangka Knowledge Discovery in Database (KDD) yang mencakup seleksi data, praproses, transformasi, penerapan algoritma data mining, dan evaluasi hasil.

2.1 Pengumpulan dan Seleksi Data

Data dikumpulkan melalui tiga metode: observasi langsung terhadap sistem pencatatan proyek, wawancara dengan manajer proyek dan tim produksi, serta dokumentasi portofolio yang sudah ada. Variabel yang digunakan sebagai atribut clustering adalah Budget Produksi (dalam juta rupiah) dan Durasi Pengerjaan (dalam hari kerja). Pemilihan dua variabel ini didasarkan

pada pertimbangan bahwa budget dan durasi merupakan dua dimensi utama yang paling berpengaruh dalam menentukan kompleksitas dan kebutuhan sumber daya suatu proyek video [5]. Total 30 proyek dari berbagai kategori berhasil dikumpulkan dengan rentang nilai budget Rp 6 juta hingga Rp 250 juta, dan durasi 2 hingga 50 hari kerja.

2.2 Normalisasi Data (Min-Max Normalization)

Sebelum proses clustering dijalankan, seluruh data dinormalisasi menggunakan metode Min-Max Normalization agar setiap variabel berada dalam rentang [0, 1]. Normalisasi ini penting karena perbedaan skala nilai yang besar antara variabel budget (ratusan juta) dan durasi (puluhan hari) dapat menyebabkan dominasi satu variabel dalam perhitungan jarak Euclidean, yang menghasilkan kluster yang bias. Formula Min-Max Normalization yang digunakan adalah:

$$x' = (x - x_{\min}) / (x_{\max} - x_{\min})$$

Untuk variabel Budget: $x_{\min} = 6$, $x_{\max} = 250$, Range = 244. Untuk variabel Durasi: $x_{\min} = 2$, $x_{\max} = 50$, Range = 48.

2.3 Penerapan Algoritma K-Means

Algoritma K-Means dijalankan dengan jumlah kluster $K=3$, yang mewakili kategorisasi Proyek Premium (budget tinggi, durasi panjang), Proyek Standard (budget dan durasi menengah), dan Proyek Quick (budget rendah, durasi singkat). Penentuan $K=3$ dilakukan berdasarkan analisis domain dan kebutuhan operasional agency. Tahapan algoritma K-Means yang diterapkan adalah: (1) Inisialisasi tiga centroid awal secara acak dari dataset; (2) Perhitungan jarak Euclidean setiap titik data ke masing-masing centroid menggunakan formula $d(x,y) = \sqrt{\sum((x_i - y_i)^2)}$; (3) Penugasan setiap proyek ke kluster dengan jarak centroid terdekat; (4) Pembaruan posisi centroid sebagai rata-rata koordinat semua anggota kluster; (5) Pengulangan langkah 2-4 hingga tidak ada perubahan penugasan kluster (konvergensi).

2.4 Evaluasi Kualitas Clustering

Kualitas hasil pengelompokan dievaluasi secara kuantitatif menggunakan dua metrik: Silhouette Score yang mengukur seberapa mirip

suatu objek dengan kluster miliknya dibandingkan kluster lain (nilai mendekati 1 menunjukkan kualitas baik), dan Davies-Bouldin Index yang mengukur rasio dispersi dalam kluster terhadap separasi antar kluster (nilai mendekati 0 menunjukkan kualitas baik). Selain evaluasi kuantitatif, dilakukan juga validasi kualitatif dengan manajer proyek berpengalaman untuk mengkonfirmasi kesesuaian hasil clustering dengan pemahaman praktis di lapangan.

3. Hasil dan Pembahasan

3.1 Data Proyek Video

Tabel 1 menyajikan sebagian dari 30 data proyek video PT Logeeka Grup ID tahun 2024 yang digunakan sebagai dataset penelitian. Data mencakup berbagai jenis produksi video dengan rentang nilai yang sangat bervariasi, mencerminkan keberagaman portofolio agensi.

Tabel 1. Sampel Data Proyek Video PT Logeeka Grup ID (10 dari 30 data)

No	Nama Proyek	Budget (Juta Rp)	Durasi (Hari)
1	Corporate Profile Bank	85	30
2	Social Media Content Batch	15	5
3	TV Commercial Product A	250	45
4	Instagram Reels Series	8	3
5	Company Event Coverage	35	10
6	YouTube Brand Video	45	14
7	Product Launch Video	120	25
8	Tutorial Video Series	20	7
9	Documentary Short Film	180	40
10	Wedding Cinematic	25	8

Sumber: Dokumen Pribadi PT Logeeka Grup ID (2024)

3.2 Normalisasi Data

Tabel 2 menampilkan hasil normalisasi Min-Max untuk 10 proyek pertama dari total 30 data.

Seluruh nilai budget dan durasi berhasil dikonversi ke dalam rentang [0, 1] sehingga kedua variabel memiliki bobot yang setara dalam proses perhitungan jarak Euclidean.

Tabel 2. Hasil Normalisasi Min-Max (10 dari 30 data)

No	Nama Proyek	Budget (Normal)	Durasi (Normal)
1	Corporate Profile Bank	0.32	0.58
2	Social Media Content Batch	0.04	0.06
3	TV Commercial Product A	1.00	0.90
4	Instagram Reels Series	0.01	0.02
5	Company Event Coverage	0.12	0.17
6	YouTube Brand Video	0.16	0.25
7	Product Launch Video	0.47	0.48
8	Tutorial Video Series	0.06	0.10
9	Documentary Short Film	0.71	0.79
10	Wedding Cinematic	0.08	0.13

Sumber: Hasil Perhitungan Peneliti (2024)

3.3 Proses Iterasi K-Means

Proses K-Means dimulai dengan inisialisasi tiga centroid awal. Pada penelitian ini, centroid awal dipilih berdasarkan representasi visual distribusi data, yakni $C1 = (0.32, 0.58)$, $C2 = (0.04, 0.06)$, dan $C3 = (1.00, 0.90)$ yang merepresentasikan titik-titik dari kelompok menengah, kecil, dan besar secara berturut-turut.

Setelah Iterasi 1, posisi centroid diperbarui berdasarkan rata-rata anggota masing-masing kluster. Komposisi kluster berubah pada beberapa titik data. Iterasi 2 menghasilkan konvergensi di mana tidak ada perpindahan keanggotaan data antar kluster, sehingga proses dihentikan. Tabel 3 menyajikan centroid akhir setelah iterasi selesai.

Tabel 3. Posisi Centroid Akhir Setelah Konvergensi

Kluster	Kategori	Centroid Budget	Centroid Durasi
---------	----------	-----------------	-----------------

K1	Proyek Standard	0.29	0.42
K2	Proyek Quick	0.03	0.05
K3	Proyek Premium	0.84	0.82

Sumber: Hasil Perhitungan Peneliti (2024)

3.4 Hasil Pengelompokan Final

Tabel 4 menampilkan hasil pengelompokan akhir seluruh 30 proyek video ke dalam tiga kluster. Kluster 3 (Proyek Premium) mencakup proyek-proyek dengan budget di atas Rp 100 juta dan durasi lebih dari 21 hari. Kluster 1 (Proyek Standard) mengelompokkan proyek dengan skala menengah. Kluster 2 (Proyek Quick) mengumpulkan proyek berbudget di bawah Rp 30 juta dengan durasi singkat.

Tabel 4. Hasil Pengelompokan Final 30 Proyek Video

No	Nama Proyek	Budget	Durasi	Kluster
1	Corporate Profile Bank	85	30	1 (Standard)
2	Social Media Content Batch	15	5	2 (Quick)
3	TV Commercial Product A	250	45	3 (Premium)
4	Instagram Reels Series	8	3	2 (Quick)
5	Company Event Coverage	35	10	1 (Standard)
6	YouTube Brand Video	45	14	1 (Standard)
7	Product Launch Video	120	25	3 (Premium)
8	Tutorial Video Series	20	7	2 (Quick)
9	Documentary Short Film	180	40	3 (Premium)
10	Wedding Cinematic	25	8	2 (Quick)

11	Music Video Production	95	21	1 (Standard)
12	Testimonial Video	12	4	2 (Quick)
13	Property Showcase	30	9	2 (Quick)
14	Fashion Campaign	150	35	3 (Premium)
15	Food & Beverage TVC	200	42	3 (Premium)
16	Explainer Animation	55	18	1 (Standard)
17	Event Highlight Reel	18	6	2 (Quick)
18	Brand Story Documentary	220	50	3 (Premium)
19	TikTok Content Pack	10	3	2 (Quick)
20	Training Video Module	40	12	1 (Standard)
21	Hotel Promotional Video	75	20	1 (Standard)
22	LinkedIn Corporate Update	14	5	2 (Quick)
23	E-commerce Product Demo	22	7	2 (Quick)
24	Festival Aftermovie	65	15	1 (Standard)
25	Charity Campaign Video	80	22	1 (Standard)
26	Instagram Story Ads	6	2	2 (Quick)
27	Corporate Training Series	110	28	3 (Premium)
28	Real Estate Virtual Tour	50	13	1 (Standard)
29	Startup Pitch Video	28	8	2 (Quick)

30	Annual Report Video	90	24	1 (Standard)
----	---------------------	----	----	------------------

Sumber: Hasil Perhitungan K-Means Peneliti (2024)

Distribusi keanggotaan kluster adalah: Kluster 1 (Proyek Standard) = 11 proyek (36,7%), Kluster 2 (Proyek Quick) = 11 proyek (36,7%), dan Kluster 3 (Proyek Premium) = 8 proyek (26,7%). Hasil ini menunjukkan bahwa portofolio PT Logeeka Grup ID didominasi oleh proyek-proyek skala kecil hingga menengah, sementara proyek premium yang membutuhkan sumber daya besar menempati sepertiga dari total pekerjaan.

3.5 Evaluasi Kualitas Clustering

Evaluasi menggunakan Silhouette Score menghasilkan nilai rata-rata sebesar 0,71 yang tergolong dalam kategori struktur pengelompokan yang kuat (strong clustering structure), mengindikasikan bahwa data dalam setiap kluster memiliki kemiripan yang tinggi satu sama lain dan berbeda secara signifikan dari data di kluster lain. Davies-Bouldin Index menghasilkan nilai 0,38, yang menunjukkan kluster-kluster yang kompak dengan separasi yang jelas. Kedua nilai ini mengkonfirmasi bahwa pemilihan $K=3$ merupakan keputusan yang tepat untuk dataset ini.

Validasi kualitatif yang dilakukan bersama manajer proyek PT Logeeka Grup ID menunjukkan bahwa kategori Proyek Premium, Standard, dan Quick selaras dengan pemahaman operasional tim produksi selama ini. Manajer proyek menegaskan bahwa pengelompokan tersebut mencerminkan perbedaan nyata dalam kebutuhan tim, peralatan, dan tahapan produksi yang diperlukan untuk setiap jenis proyek.

3.6 Implementasi Sistem Web

Sistem aplikasi dibangun menggunakan bahasa pemrograman PHP dengan framework berbasis MVC sederhana dan basis data MySQL. Antarmuka pengguna dirancang responsif menggunakan Bootstrap 5 untuk memastikan aksesibilitas dari berbagai perangkat. Sistem terdiri dari delapan modul utama: Autentikasi Login, Kelola Data Client, Kelola Proyek Video, Kelola Tim Produksi, Kelola Budget, Proses

Clustering K-Means, Visualisasi Hasil, dan Laporan Cetak.

Modul Clustering merupakan inti dari sistem, di mana pengguna dapat memicu proses K-Means secara langsung melalui antarmuka web. Sistem secara otomatis mengambil data proyek dari database, menjalankan proses normalisasi dan iterasi K-Means, lalu menyimpan dan menampilkan hasil pengelompokan beserta visualisasi posisi centroid. Modul Laporan memungkinkan manajer proyek mengunduh hasil clustering dalam format yang dapat langsung digunakan untuk perencanaan alokasi sumber daya.

Pengujian sistem dilakukan menggunakan metode Black-Box Testing pada seluruh fungsionalitas dengan 45 skenario uji. Hasil pengujian menunjukkan tingkat keberhasilan 100% di mana semua fitur berjalan sesuai dengan spesifikasi yang telah ditetapkan, termasuk validasi input data, proses perhitungan algoritma, dan tampilan output laporan.

4. Kesimpulan dan Saran

Penelitian ini berhasil mewujudkan sistem aplikasi web berbasis K-Means Clustering yang mampu mengelompokkan 30 proyek video PT Logeeka Grup ID ke dalam tiga kluster bermakna secara otomatis. Proses iterasi K-Means mencapai konvergensi pada iterasi ke-2 dengan hasil pengelompokan: 11 proyek masuk Kluster Standard (36,7%), 11 proyek masuk Kluster Quick (36,7%), dan 8 proyek masuk Kluster Premium (26,7%). Evaluasi kuantitatif menghasilkan Silhouette Score = 0,71 dan Davies-Bouldin Index = 0,38, keduanya mengindikasikan kualitas clustering yang kuat dan valid.

Penerapan sistem terbukti mampu mengatasi ketergantungan pada penilaian subjektif dalam proses kategorisasi proyek, menyediakan basis data klasifikasi yang konsisten untuk pengambilan keputusan alokasi tim produksi, dan mempercepat proses estimasi anggaran berdasarkan pola kluster historis. Hal ini sejalan dengan temuan Firmansyah (2023) yang mencatat peningkatan akurasi estimasi dari 52% menjadi 84% setelah implementasi sistem berbasis data di lingkungan agency kreatif [6].

Untuk pengembangan lebih lanjut, disarankan: (1) memperluas variabel atribut clustering dengan menambahkan jumlah revisi, ukuran tim produksi, dan tingkat kerumitan teknis; (2) mengeksplorasi algoritma clustering lanjutan seperti DBSCAN atau Gaussian Mixture Model untuk menangani distribusi data yang tidak berbentuk bola; (3) mengintegrasikan sistem dengan modul penjadwalan proyek dan manajemen sumber daya manusia yang lebih komprehensif; serta (4) mengembangkan model prediktif berbasis supervised learning untuk memperkirakan kluster proyek baru secara real-time sejak tahap negosiasi kontrak.

DAFTAR PUSTAKA

- [1] Suryadi, A., & Mulyana, D. (2021). Analisis Permasalahan Manajemen Proyek pada Agency Kreatif Digital di Indonesia. *Jurnal Manajemen Teknologi Informasi*, 12(2), 45-58.
- [2] Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann Publishers.
- [3] Kusumawati, R., & Wibowo, A. (2022). Segmentasi Pelanggan Platform E-Commerce Menggunakan Algoritma K-Means Clustering. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(1), 112-121.
- [4] Andini, T. D., & Prasetya, D. A. (2021). Optimasi Jumlah Kluster pada K-Means Menggunakan Elbow Method dan Silhouette Score. *Jurnal Informatika*, 15(3), 78-89.
- [5] Rahman, F., & Hidayat, R. (2023). Pengaruh Normalisasi Data terhadap Akurasi Algoritma K-Means pada Data Proyek Kreatif. *Jurnal Ilmu Komputer dan Informasi*, 16(1), 23-35.
- [6] Firmansyah, B. (2023). Implementasi Data-Driven Decision Making pada Agency Produksi Video: Studi Kasus PT Creativindo Media. *Jurnal Bisnis dan Teknologi Informasi*, 8(2), 201-215.
- [7] Xu, R., & Tian, Y. (2017). A Comprehensive Study of Clustering Algorithms. *Annals of Data Science*, 2(2), 165-193.
- [8] MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281-297.
- [9] Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., ... & Lin, C. T. (2017). A Review of Clustering Techniques and Developments. *Neurocomputing*, 267, 664-681.
- [10] Putri, S. N., & Nugroho, H. (2023). Tantangan Adopsi Sistem Berbasis Data pada Tim Kreatif di Industri Periklanan Digital Indonesia. *Jurnal Komunikasi dan Bisnis*, 11(1), 88-103.