E-ISSN: 3025-0994
Halaman 5779-5790
Volume 4 Nomor 3 Tahun 2026

Pemetaan Risiko Pembiayaan Konsumen Berbasis Clustering Wilayah Pengajuan

Muchammad Ubaidillah[1], Yitno Utomo[2]

[1,2] Program Studi Teknik Industri, Fakultas Teknik dan Sains, Universitas PGRI Adi Buana Surabaya, Indonesia
e-mail: ubaidillahst0@gmail.com ; yitno@unipasby.ac.id

Abstrak

Risiko gagal bayar pada perusahaan pembiayaan tidak selalu tersebar secara merata karena karakteristik pembayaran konsumen dapat berbeda antarwilayah. Penelitian ini bertujuan memetakan tingkat risiko pembiayaan berdasarkan wilayah pengajuan konsumen serta merumuskan strategi pengendalian yang sesuai dengan profil setiap kelompok. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan studi kasus pada perusahaan pembiayaan di Sidoarjo. Data sekunder mencakup 712 konsumen yang tersebar pada 40 wilayah selama 2025 dan dikelompokkan berdasarkan proporsi pembayaran tepat waktu, keterlambatan 1–90 hari, serta keterlambatan 91–210 hari. Data dibersihkan, ditransformasi menjadi proporsi, kemudian dianalisis menggunakan metode Elbow, algoritma K-Means berbasis jarak Euclidean, dan evaluasi deskriptif antarcluster. Hasil menunjukkan bahwa tiga cluster merupakan struktur yang paling informatif. Cluster 1 terdiri atas 13 wilayah dengan profil WARNING, cluster 2 mencakup 19 wilayah dengan profil EXCELLENT, dan cluster 3 berisi 8 wilayah dengan profil BAD. Centroid cluster berisiko tinggi memperlihatkan proporsi pembayaran tepat waktu 0,6901, keterlambatan 1–90 hari 0,2661, dan keterlambatan 91–210 hari 0,0438. Wilayah Jabon menunjukkan proporsi keterlambatan panjang tertinggi. Penelitian menyimpulkan bahwa clustering wilayah dapat menyediakan dasar objektif untuk diferensiasi kebijakan pembiayaan. Implikasinya, perusahaan perlu menerapkan pemantauan preventif pada wilayah WARNING, ekspansi terukur pada wilayah EXCELLENT, serta pengetatan persetujuan, kunjungan lapangan, dan restrukturisasi pada wilayah BAD.

Kata Kunci: K-Means; Risiko Pembiayaan; Segmentasi Wilayah; Keterlambatan Pembayaran; Data Mining

Abstract

Credit default risk in financing companies is not necessarily distributed evenly because consumer repayment behavior may differ across application areas. This study aimed to map financing risk by consumer application area and formulate control strategies that match the profile of each group. A descriptive quantitative case study was conducted at a financing company in Sidoarjo. The secondary dataset covered 712 consumers from 40 areas during

2025 and was represented by the proportions of on-time payment, delinquency of 1–90 days, and delinquency of 91–210 days. The data were cleaned, transformed into comparable proportions, and analyzed using the Elbow method, Euclidean-distance K-Means, and descriptive evaluation of cluster differences. The results show that a three-cluster structure is the most informative. Cluster 1 contains 13 areas with a WARNING profile, Cluster 2 comprises 19 areas with an EXCELLENT profile, and Cluster 3 consists of 8 areas with a BAD profile. The high-risk centroid records an on-time payment proportion of 0.6901, a 1–90-day delinquency proportion of 0.2661, and a 91–210-day delinquency proportion of 0.0438. Jabon has the highest long-delinquency proportion. The study concludes that regional clustering provides an objective basis for differentiated financing policies. Financing companies should therefore apply preventive monitoring to WARNING areas, controlled expansion to EXCELLENT areas, and stricter approval, field visits, and restructuring to BAD areas.

Keywords: Credit Risk; Data Mining; K-Means; Payment Delinquency; Regional Segmentation

LATAR BELAKANG

Pertumbuhan industri pembiayaan memperluas akses masyarakat terhadap kendaraan, modal kerja, dan kebutuhan konsumtif, tetapi pada saat yang sama meningkatkan eksposur perusahaan terhadap keterlambatan pembayaran dan gagal bayar. Risiko tersebut tidak hanya ditentukan oleh karakter individu pemohon, melainkan juga oleh konteks wilayah, kedalaman basis pelanggan, pola ekonomi lokal, dan efektivitas penagihan. Pengelolaan risiko yang hanya bergantung pada pemeriksaan manual cenderung menghasilkan keputusan yang seragam untuk wilayah dengan karakter berbeda. Literatur mutakhir menunjukkan bahwa analitik berbasis data mampu meningkatkan ketepatan penilaian risiko, terutama ketika heterogenitas debitur dipetakan terlebih dahulu sebelum kebijakan pembiayaan ditetapkan (Caruso et al., 2021; Gavira-Durón et al., 2021; Shi et al., 2022).

Clustering menawarkan pendekatan eksploratif untuk membentuk kelompok yang relatif homogen tanpa membutuhkan label risiko yang ditentukan sebelumnya. Dalam konteks kredit, pembentukan segmen dapat membantu lembaga keuangan mengidentifikasi kelompok berisiko rendah, menengah, dan tinggi, kemudian menyesuaikan persetujuan, pemantauan, serta penagihan. Pendekatan berbasis cluster juga relevan ketika distribusi pelanggan tidak seimbang dan informasi wilayah lebih mudah diperoleh daripada variabel individual yang lengkap. Sejumlah penelitian menemukan bahwa pemisahan sampel sebelum pemodelan atau pengambilan keputusan dapat meningkatkan interpretabilitas dan kinerja analisis risiko (Yuan et al., 2022; Machado & Karray, 2022b; Baser et al., 2023; Bosker et al., 2025).

Perusahaan pembiayaan yang menjadi objek penelitian melayani pengajuan dari 40 wilayah dengan total 712 konsumen pada periode 2025. Data pembayaran memperlihatkan 562 konsumen membayar tepat waktu, 135 konsumen mengalami keterlambatan 1–90 hari, dan 15 konsumen mengalami keterlambatan 91–210 hari. Angka agregat tersebut menggambarkan portofolio yang secara umum sehat, tetapi dapat menutupi konsentrasi risiko pada wilayah tertentu. Keterlambatan panjang merupakan sinyal penting karena berkaitan

dengan kebutuhan restrukturisasi, intensifikasi penagihan, dan potensi penurunan kualitas aset. Oleh sebab itu, proporsi status pembayaran per wilayah perlu dianalisis sebagai pola, bukan hanya dijumlahkan pada tingkat perusahaan.

Penelitian terdahulu di Indonesia telah menggunakan K-Means untuk segmentasi nasabah kredit dan pelanggan pembiayaan, tetapi sebagian besar berfokus pada demografi, nilai transaksi, atau perilaku individual (Ahsina et al., 2022; Akbar et al., 2023; Rizkyanto & Gaol, 2023; Ocktavia & Atmojo, 2025). Penelitian internasional juga menunjukkan manfaat pengelompokan dalam pembentukan rating rumah tangga, estimasi probabilitas gagal bayar, evaluasi risiko campuran, dan pemulihan kredit bermasalah (Dang et al., 2022; Carleo & Rocci, 2024; Jamotton et al., 2024). Namun, kajian yang secara khusus membangun pemetaan risiko berdasarkan wilayah pengajuan dengan tiga tingkat delinquency pada perusahaan pembiayaan lokal masih terbatas. Kesenjangan lainnya terletak pada kecenderungan penelitian untuk menerima keluaran perangkat lunak tanpa membahas ketimpangan ukuran sampel wilayah dan batas interpretasi ANOVA pada K-Means.

Berdasarkan kesenjangan tersebut, penelitian ini menjawab tiga pertanyaan: bagaimana sebaran pembayaran tepat waktu dan keterlambatan pada setiap wilayah; bagaimana K-Means mengelompokkan wilayah berdasarkan profil tersebut; dan bagaimana hasil cluster dapat diterjemahkan menjadi strategi pembiayaan. Tujuan penelitian adalah menentukan jumlah cluster yang informatif, mengidentifikasi karakter setiap kelompok, menilai pemisahan antarcluster secara deskriptif, serta merumuskan kebijakan diferensial. Kontribusi ilmiahnya berupa kerangka segmentasi risiko wilayah yang menggabungkan normalisasi proporsi, Elbow Method, interpretasi centroid, dan kehati-hatian terhadap ukuran sampel. Kontribusi praktisnya adalah menyediakan peta prioritas yang dapat digunakan untuk persetujuan pembiayaan, early warning, penagihan, dan ekspansi pasar.

METODE

Penelitian menggunakan pendekatan kuantitatif deskriptif dengan desain studi kasus pada sebuah perusahaan pembiayaan yang melayani konsumen di Sidoarjo dan wilayah sekitarnya. Populasi penelitian adalah seluruh wilayah penanganan pembiayaan pada periode 2025. Sampel menggunakan sensus terhadap 40 wilayah yang memiliki catatan pembayaran dan dapat dipetakan secara konsisten. Unit analisis adalah wilayah, bukan individu, sehingga setiap observasi mewakili profil agregat konsumen pada satu wilayah. Data mencakup 712 konsumen dengan tiga variabel input: proporsi NO OD atau pembayaran tepat waktu, proporsi OD 1–90 hari, dan proporsi OD 91–210 hari.

Data diperoleh dari catatan internal perusahaan dan dokumentasi status pembayaran. Tahap prapemrosesan meliputi pemeriksaan kelengkapan, penyamaan nama wilayah, agregasi jumlah konsumen menurut kategori, dan transformasi jumlah menjadi proporsi terhadap total konsumen pada setiap wilayah. Transformasi proporsi diperlukan agar wilayah berukuran besar dan kecil dapat dibandingkan pada skala yang sama. Data kemudian dianalisis menggunakan IBM SPSS Statistics. Jumlah cluster diuji pada K=2 sampai K=6 melalui perubahan Sum of Squared Errors, sedangkan K-Means menggunakan jarak Euclidean dan centroid awal yang dipilih oleh prosedur perangkat lunak.

Proses iterasi dihentikan setelah centroid tidak lagi berubah. Hasil dianalisis melalui final cluster centers, jarak antarcentroid, jumlah dan anggota wilayah, serta profil risiko yang

diberi label EXCELLENT, WARNING, dan BAD. Tabel ANOVA keluaran K-Means digunakan secara deskriptif untuk mengidentifikasi variabel yang paling membedakan cluster, bukan sebagai uji hipotesis independen, karena cluster dibentuk untuk memaksimalkan perbedaan antarobjek. Interpretasi juga mempertimbangkan ukuran basis konsumen per wilayah agar cluster dengan pembayaran sempurna tetapi sampel sangat kecil tidak langsung dianggap bebas risiko.

HASIL DAN PEMBAHASAN

Profil Portofolio dan Kualitas Basis Data

Portofolio penelitian terdiri atas 712 konsumen pada 40 wilayah. Sebanyak 78,93% konsumen berada pada kategori NO OD, 18,96% berada pada OD 1–90 hari, dan 2,11% berada pada OD 91–210 hari. Komposisi tersebut menunjukkan bahwa mayoritas portofolio membayar tepat waktu, tetapi 150 konsumen telah memasuki status terlambat. Secara manajerial, keterlambatan pendek dan panjang tidak dapat diperlakukan sama. OD 1–90 hari masih membuka ruang intervensi preventif melalui pengingat dan kunjungan, sedangkan OD 91–210 hari membutuhkan penanganan yang lebih intensif karena menunjukkan tekanan kemampuan bayar yang lebih serius.

Tabel 1. Komposisi status pembayaran konsumen

Status pembayaran	Jumlah konsumen	Proporsi	Makna risiko
NO OD	562	78,93%	Pembayaran tepat waktu
OD 1–90 hari	135	18,96%	Early warning dan kebutuhan pemantauan
OD 91–210 hari	15	2,11%	Risiko tinggi dan potensi gagal bayar
Total	712	100,00%	Basis analisis 40 wilayah

Sumber: Data penelitian diolah (2026).

Penggunaan proporsi sebagai variabel input mengurangi bias yang timbul akibat perbedaan jumlah konsumen antarwilayah. Tanpa transformasi, wilayah dengan volume pelanggan besar cenderung mendominasi jarak Euclidean walaupun tingkat risikonya tidak lebih tinggi. Pilihan ini sejalan dengan prinsip bahwa clustering harus membandingkan pola relatif pada skala yang sepadan. Meskipun demikian, proporsi juga memiliki kelemahan: nilai 100% pembayaran tepat waktu pada wilayah dengan satu pelanggan tidak memiliki tingkat kepastian yang sama dengan nilai serupa pada wilayah dengan puluhan pelanggan. Oleh karena itu, hasil cluster perlu dipadukan dengan jumlah eksposur sebelum digunakan sebagai keputusan final. Literatur clustered credit modeling menekankan bahwa kualitas segmentasi sangat tergantung pada pemilihan dan representasi variabel (Caruso et al., 2021; Bosker et al., 2025).

Penentuan Jumlah Cluster dan Konvergensi

Elbow Method memperlihatkan penurunan total SSE dari 0,266 pada K=2 menjadi 0,111 pada K=3, atau berkurang 0,155. Setelah titik tersebut, penurunan melambat menjadi

0,039 pada K=4, 0,037 pada K=5, dan 0,035 pada K=6. Pola tersebut menunjukkan bahwa tiga cluster memberikan keseimbangan antara homogenitas internal dan kesederhanaan interpretasi. Pemilihan K=3 juga memiliki relevansi kebijakan karena membentuk tiga tingkat tindakan yang mudah diterapkan: ekspansi terkendali, pemantauan preventif, dan penanganan intensif. Penentuan jumlah cluster tidak seharusnya hanya mengikuti nilai SSE terendah, karena SSE selalu menurun ketika jumlah cluster bertambah; yang dicari adalah titik tambahan manfaat yang mulai mengecil.

Tabel 2. Hasil Elbow Method

K	SSE NO OD	SSE OD 1–90	SSE OD 91–210	Total SSE	Penurunan
2	0,114	0,114	0,038	0,266	–
3	0,037	0,037	0,037	0,111	0,155
4	0,036	0,036	0,000	0,072	0,039
5	0,035	0,000	0,000	0,035	0,037
6	0,000	0,000	0,000	0,000	0,035

Sumber: Data penelitian diolah (2026).

Algoritma mencapai konvergensi pada iterasi kelima, ketika seluruh perubahan centroid menjadi 0,000. Konvergensi menunjukkan bahwa keanggotaan wilayah telah stabil terhadap centroid yang terbentuk, tetapi tidak otomatis membuktikan bahwa solusi merupakan optimum global. K-Means sensitif terhadap inisialisasi awal, sehingga replikasi dengan beberapa random starts atau pembandingan dengan hierarchical clustering dan K-medoids akan memperkuat validitas. Temuan ini relevan dengan penelitian Yuan et al. (2022) dan Baser et al. (2023), yang menunjukkan bahwa penentuan struktur cluster dan pemilihan variabel merupakan tahap kritis sebelum cluster digunakan dalam sistem prediksi risiko.

Karakteristik Tiga Cluster Risiko

Hasil akhir membentuk Cluster 1 sebanyak 13 wilayah, Cluster 2 sebanyak 19 wilayah, dan Cluster 3 sebanyak 8 wilayah. Cluster 1 memiliki centroid NO OD 0,8304, OD 1–90 hari 0,1535, dan OD 91–210 hari 0,0161. Profil ini dikategorikan WARNING karena sebagian besar pembayaran masih tepat waktu, tetapi keterlambatan telah muncul pada dua horizon. Cluster 2 mempunyai centroid 1,0000; 0,0000; dan 0,0000, sehingga dikategorikan EXCELLENT. Cluster 3 memiliki centroid 0,6901; 0,2661; dan 0,0438, sehingga dikategorikan BAD. Jarak terbesar terjadi antara Cluster 2 dan Cluster 3 sebesar 0,411, yang menegaskan perbedaan kuat antara kelompok pembayaran sempurna dan kelompok keterlambatan tinggi.

Tabel 3. Ringkasan karakteristik cluster

Cluster	Kategori	Jumlah wilayah	NO OD	OD 1–90	OD 91–210	Interpretasi
1	WARNING	13	0,8304	0,1535	0,0161	Risiko sedang; membutuhkan pencegahan

Cluster	Kategori	Jumlah wilayah	NO OD	OD 1–90	OD 91–210	Interpretasi
2	EXCELLENT	19	1,0000	0,0000	0,0000	Risiko tercatat sangat rendah
3	BAD	8	0,6901	0,2661	0,0438	Risiko tinggi; prioritas mitigasi

Sumber: Data penelitian diolah (2026).

Cluster WARNING mencakup Balongbendo, Buduran, Driyorejo, Gedangan, Sambikerep, Sedati, Sidoarjo, Sukodono, Taman, Tanggulangin, Tarik, Tulangan, dan Wonoayu. Tarik berada paling dekat dengan centroid, sehingga paling merepresentasikan profil rata-rata kelompok. Cluster ini bukan kelompok gagal bayar, melainkan ruang intervensi dini. Proporsi keterlambatan pendek yang masih terlihat memberi peluang bagi perusahaan untuk mencegah migrasi menuju delinquency panjang melalui reminder sebelum jatuh tempo, kunjungan selektif, dan pemantauan tren.

Cluster EXCELLENT berisi Dlanggu, Gayungan, Gempol, Gubeng, Jambangan, Karangpilang, Kedamean, Krembangan, Lakar Santri, Mulyorejo, Pungging, Sawahan, Semampir, Simokerto, Sukolilo, Wiyung, Wonocolo, Wonokromo, dan Wringin Anom. Semua wilayah tercatat memiliki proporsi pembayaran tepat waktu 1,0000. Namun, sebagian besar wilayah ini memiliki basis konsumen kecil dan beberapa berada di luar Kabupaten Sidoarjo. Karena itu, label EXCELLENT harus dibaca sebagai performa observasi, bukan jaminan risiko nol. Dalam manajemen portofolio, ekspansi pada kelompok ini sebaiknya dilakukan bertahap dengan batas eksposur dan pemantauan vintage agar kualitas cluster tidak berubah ketika volume meningkat.

Cluster BAD terdiri atas Candi, Jabon, Krembung, Krian, Porong, Prambon, Rungkut, dan Waru. Jabon memiliki proporsi OD 91–210 hari sebesar 0,1667, tertinggi dalam data, sedangkan Porong sebesar 0,1000. Keterlambatan panjang pada dua wilayah tersebut mengindikasikan masalah yang tidak dapat diselesaikan dengan reminder standar. Temuan ini konsisten dengan kajian clustering kredit yang menyatakan bahwa kelompok berisiko tinggi memerlukan kebijakan terpisah, pengawasan lebih ketat, dan strategi remedial yang spesifik (Gavira-Durón et al., 2021; Barragán Garnica, 2023; Yu et al., 2025).

Pemisahan Cluster dan Makna Statistik

Output ANOVA K-Means menunjukkan nilai F deskriptif 437,541 untuk NO OD, 203,619 untuk OD 1–90 hari, dan 6,771 untuk OD 91–210 hari. Nilai tersebut menunjukkan bahwa proporsi pembayaran tepat waktu merupakan variabel yang paling kuat membedakan cluster, diikuti keterlambatan pendek. Keterlambatan panjang tetap membedakan kelompok, tetapi kontribusinya lebih kecil karena jumlah kasusnya rendah. Secara substantif, temuan tersebut berarti identitas cluster terutama ditentukan oleh kemampuan wilayah mempertahankan pembayaran normal, sedangkan OD 91–210 berfungsi sebagai sinyal intensitas risiko pada kelompok tertentu.

Tabel 4. Evaluasi deskriptif perbedaan antarcluster

Variabel	F deskriptif	Sig. keluaran	Makna analitis
NO OD	437,541	0,000	Pembeda utama antarcluster
OD 1–90 hari	203,619	0,000	Pembeda kuat dan indikator early warning
OD 91–210 hari	6,771	0,003	Pembeda risiko berat dengan frekuensi rendah

Sumber: Data penelitian diolah (2026).

Nilai signifikansi pada tabel tersebut tidak boleh diperlakukan sebagai bukti inferensial yang independen karena K-Means membentuk cluster dengan tujuan memaksimalkan perbedaan. Dengan demikian, ANOVA berfungsi untuk membaca variabel pembeda, bukan menguji hipotesis bahwa cluster dari populasi acak memiliki rata-rata berbeda. Pendekatan yang lebih kuat pada penelitian berikutnya adalah menggunakan validasi internal seperti silhouette coefficient dan Davies–Bouldin Index, melakukan stability testing pada beberapa inisialisasi, serta menilai validitas eksternal menggunakan status gagal bayar pada periode setelah pembentukan cluster. Kehati-hatian ini penting agar keluaran statistik tidak menghasilkan keyakinan berlebihan.

Implikasi Kebijakan dan Kontribusi Akademik

Segmentasi wilayah mengubah pengelolaan risiko dari kebijakan seragam menjadi alokasi sumber daya berbasis prioritas. Pada Cluster WARNING, kebijakan yang relevan adalah early warning, reminder sebelum jatuh tempo, dan kunjungan selektif pada konsumen yang mulai menunggak. Pada Cluster EXCELLENT, perusahaan dapat mempertahankan layanan dan melakukan ekspansi terukur, tetapi tetap menggunakan batas eksposur karena beberapa wilayah didukung oleh sampel kecil. Pada Cluster BAD, proses analisis calon konsumen perlu diperketat, persetujuan dapat dinaikkan ke level manajemen, dan penanganan existing customer diarahkan pada kunjungan intensif, restrukturisasi, atau recovery sesuai kemampuan bayar.

Tabel 5. Strategi pengendalian risiko berdasarkan cluster

Cluster	Fokus kebijakan	Tindakan prioritas	Indikator pemantauan
WARNING	Pencegahan migrasi risiko	Reminder dini, kunjungan selektif, review kelayakan wilayah	Rasio OD 1–90 dan perpindahan ke OD 91–210
EXCELLENT	Pertahankan kualitas dan ekspansi terukur	Program loyalitas, percepatan proses, limit eksposur bertahap	Vintage delinquency dan pertumbuhan konsumen

Cluster	Fokus kebijakan	Tindakan prioritas	Indikator pemantauan
BAD	Mitigasi dan pemulihan	Approval tambahan, penagihan intensif, restrukturisasi, red-area control	Cure rate, roll rate, recovery, dan NPF/NPL wilayah

Sumber: Data penelitian diolah (2026).

Kontribusi akademik penelitian ini terletak pada pemanfaatan status pembayaran sebagai komposisi wilayah, sehingga cluster merepresentasikan struktur risiko relatif dan bukan sekadar volume pelanggan. Hasil juga memperlihatkan pentingnya membedakan interpretasi cluster dari prediksi individual. Wilayah BAD tidak berarti seluruh konsumennya buruk, dan wilayah EXCELLENT tidak berarti setiap calon konsumen aman. Clustering menyediakan prior probability untuk alokasi perhatian, sedangkan keputusan pembiayaan tetap memerlukan penilaian individual. Perspektif tersebut sejalan dengan clustered credit risk modeling yang menggabungkan segmentasi dengan model lanjutan agar heterogenitas tidak hilang (Machado & Karray, 2022b; Baser et al., 2023; Bosker et al., 2025).

Tata Kelola Implementasi dan Validasi Berkala

Peta risiko wilayah hanya bernilai apabila diintegrasikan ke dalam proses keputusan dan diperbarui secara periodik. Tahap pertama adalah menetapkan definisi operasional status pembayaran, identitas wilayah, tanggal observasi, dan nilai eksposur yang konsisten antara fungsi pemasaran, kredit, penagihan, serta manajemen risiko. Rekonsiliasi bulanan diperlukan untuk mencegah perbedaan denominator dan memastikan perpindahan konsumen antarstatus dapat ditelusuri. Praktik ini penting karena model risiko yang akurat secara teknis tetap dapat menghasilkan keputusan buruk ketika data input tidak lengkap atau proses pemantauannya tidak konsisten (Alonso-Robisco & Carbó, 2022; Bazzana et al., 2024).

Tahap kedua adalah menerjemahkan cluster menjadi aturan bantu keputusan, bukan keputusan otomatis. Label wilayah dapat digunakan sebagai prior informasi untuk menentukan intensitas verifikasi, tetapi persetujuan akhir tetap mempertimbangkan karakter, kapasitas, modal, agunan, kondisi usaha, dan riwayat individu. Pendekatan ini menghindari diskriminasi geografis dan ecological fallacy, yaitu kesalahan menyimpulkan karakter seseorang hanya dari rata-rata wilayahnya. Untuk Cluster BAD, misalnya, perusahaan dapat menambah level persetujuan dan verifikasi lapangan tanpa menutup akses pembiayaan bagi pemohon yang layak.

Tabel 6. Peta jalan penerapan clustering risiko wilayah

Tahap	Aktivitas utama	Frekuensi	Keluaran pengendalian
Data governance	Validasi identitas wilayah, status pembayaran, eksposur, dan rekonsiliasi lintas fungsi	Bulanan	Dataset terverifikasi dan audit trail

Tahap	Aktivitas utama	Frekuensi	Keluaran pengendalian
Cluster refresh	Transformasi proporsi, pengujian K, pembentukan ulang cluster, dan stability check	Triwulanan/semesteran	Peta risiko dan perpindahan cluster
Policy execution	Penerapan diferensiasi verifikasi, approval, reminder, kunjungan, dan restrukturisasi	Setiap pengajuan dan siklus tagihan	Tindakan sesuai profil risiko
Outcome validation	Bandingkan default, roll rate, cure rate, recovery, pendapatan, dan biaya	Triwulanan	Evaluasi manfaat dan kalibrasi kebijakan
Governance review	Tinjauan fairness, override manusia, keluhan, dan persetujuan manajemen	Semesteran	Model risk report dan keputusan tindak lanjut

Sumber: Sintesis penulis berdasarkan hasil penelitian (2026).

Tahap ketiga adalah melakukan pembaruan cluster secara triwulanan atau semesteran, disesuaikan dengan volume transaksi. Dashboard perlu menampilkan jumlah konsumen, outstanding, proporsi delinquency, roll rate antarstatus, cure rate, recovery, dan perubahan keanggotaan wilayah. Pergeseran wilayah dari EXCELLENT ke WARNING harus memicu investigasi dini, sedangkan perpindahan dari BAD ke WARNING perlu dikonfirmasi melalui perbaikan berkelanjutan, bukan hanya fluktuasi sampel. Studi regional pinjaman daring di Indonesia memperlihatkan bahwa outstanding dan tingkat gagal bayar dapat menghasilkan pola wilayah yang berbeda, sehingga segmentasi harus dibaca bersama ukuran eksposur (Wewengkang & Wibowo, 2026).

Tahap keempat adalah validasi hasil kebijakan. Perusahaan perlu membandingkan tingkat gagal bayar, pendapatan berbasis risiko, biaya penagihan, dan tingkat penerimaan sebelum dan sesudah cluster digunakan. Validasi out-of-sample dapat menilai apakah wilayah yang dilabeli BAD benar-benar menunjukkan default lebih tinggi pada periode berikutnya. Selain itu, analisis risiko-adjusted revenue diperlukan agar pengetatan tidak menurunkan portofolio sehat secara berlebihan (Machado & Karray, 2022a). Dengan mekanisme tersebut, clustering berfungsi sebagai sistem pembelajaran yang terus dikalibrasi, bukan klasifikasi statis yang berlaku tanpa batas waktu.

Penelitian ini memiliki keterbatasan. Pertama, variabel hanya berupa tiga proporsi status pembayaran dan belum memasukkan jumlah pembiayaan, tenor, pendapatan, jenis pekerjaan, nilai agunan, atau karakteristik ekonomi wilayah. Kedua, periode satu tahun belum menangkap perubahan siklus dan migrasi cluster. Ketiga, wilayah dengan jumlah konsumen sangat kecil dapat menghasilkan proporsi ekstrem. Keempat, K-Means mengasumsikan bentuk cluster yang relatif kompak dan sensitif terhadap centroid awal. Penelitian selanjutnya perlu menggunakan bobot berdasarkan eksposur, minimum sample threshold, data

longitudinal, dan perbandingan dengan hierarchical clustering, K-medoids, fuzzy clustering, atau model hybrid. Kajian terbaru menunjukkan bahwa penggabungan clustering dengan klasifikasi, variable selection, dan validasi out-of-sample dapat meningkatkan daya guna untuk keputusan kredit (Yuan et al., 2022; Baser et al., 2023; Bosker et al., 2025; Yu et al., 2025).

SIMPULAN

K-Means menghasilkan tiga profil risiko pada 40 wilayah: WARNING sebanyak 13 wilayah, EXCELLENT 19 wilayah, dan BAD 8 wilayah. Cluster BAD memiliki pembayaran tepat waktu terendah serta keterlambatan tertinggi, sedangkan Jabon menunjukkan proporsi OD 91–210 hari paling besar. Hasil ini menjawab bahwa segmentasi wilayah dapat menyediakan dasar objektif untuk membedakan kebijakan pembiayaan.

Perusahaan disarankan menerapkan early warning pada WARNING, ekspansi terukur pada EXCELLENT, dan pengetatan serta restrukturisasi pada BAD. Keterbatasan variabel, periode, dan ukuran sampel wilayah menuntut penelitian longitudinal dengan indikator tambahan, validasi cluster, dan perbandingan algoritma.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Universitas PGRI Adi Buana Surabaya dan perusahaan pembiayaan yang telah memberikan akses data serta dukungan selama penelitian. Artikel ini dikembangkan dari tugas akhir Program Studi Teknik Industri.

DAFTAR PUSTAKA

Ahsina, N. H., Fatimah, F., & Rachmawati, F. 2022. Customer Segmentation Analysis Based On Credit Using K-Means Clustering. Jurnal Ilmiah Teknologi Informasi Terapan, 8(3). Bandung: Universitas Widyatama. Https://Doi.Org/10.33197/Jitter.Vol8.Iss3.2022.883.

Akbar, M. N., Salsabila, A., Asri, A. P., & Syawir, M. 2023. Clustering Analysis For Credit Card User Segmentation Using K-Means Algorithm And Principal Component Analysis. AGENTS: Journal Of Artificial Intelligence And Data Science, 3(1). Makassar: UIN Alauddin. Https://Doi.Org/10.24252/Jagti.V3i1.56.

Alonso-Robisco, A., & Carbó, J. M. 2022. Measuring The Model Risk-Adjusted Performance Of Machine Learning Algorithms In Credit Default Prediction. Financial Innovation, 8, 70. Berlin: Springer Nature. Https://Doi.Org/10.1186/S40854-022-00366-1.

Barragán Garnica, D. 2023. Patrones De Comportamiento De Clientes Con Tarjetas De Crédito Con Deterioro De Calificación Por Riesgo Utilizando K-Means. ODEON, 22, 7–37. Bogotá: Universidad Externado De Colombia. Https://Doi.Org/10.18601/17941113.N22.02.

Baser, F., Koc, O., & Selcuk-Kestel, A. S. 2023. Credit Risk Evaluation Using Clustering Based Fuzzy Classification Method. Expert Systems With Applications, 223, 119882. Amsterdam: Elsevier. Https://Doi.Org/10.1016/J.Eswa.2023.119882.

Bazzana, F., Bee, M., & Khatir, A. A. H. A. 2024. Machine Learning Techniques For Default Prediction: An Application To Small Italian Companies. Risk Management, 26, 1. London: Palgrave Macmillan. Https://Doi.Org/10.1057/S41283-023-00132-2.

Bosker, J., Gürtler, M., & Zöllner, M. 2025. Machine Learning-Based Variable Selection For Clustered Credit Risk Modeling. Journal Of Business Economics, 95(4), 617–652. Berlin: Springer. Https://Doi.Org/10.1007/S11573-024-01213-8.

Carleo, A., & Rocci, R. 2024. Functional Clustering Of Npls Recovery Curves. Socio-Economic Planning Sciences, 95, 102018. Amsterdam: Elsevier. Https://Doi.Org/10.1016/J.Seps.2024.102018.

Caruso, G., Gattone, S. A., Fortuna, F., & Di Battista, T. 2021. Cluster Analysis For Mixed Data: An Application To Credit Risk Evaluation. Socio-Economic Planning Sciences, 73, 100850. Amsterdam: Elsevier. Https://Doi.Org/10.1016/J.Seps.2020.100850.

Dang, C., Chen, X., Yu, S., Chen, R., & Yang, Y. 2022. Credit Ratings Of Chinese Households Using Factor Scores And K-Means Clustering Method. International Review Of Economics & Finance, 78, 309–320. Amsterdam: Elsevier. Https://Doi.Org/10.1016/J.Iref.2021.12.014.

Gavira-Durón, N., Gutierrez-Vargas, O., & Cruz-Aké, S. 2021. Markov Chain K-Means Cluster Models And Their Use For Companies' Credit Quality And Default Probability Estimation. Mathematics, 9(8), 879. Basel: MDPI. Https://Doi.Org/10.3390/Math9080879.

Jamotton, C., Hainaut, D., & Hames, T. 2024. Insurance Analytics With Clustering Techniques. Risks, 12(9), 141. Basel: MDPI. Https://Doi.Org/10.3390/Risks12090141.

Machado, M. R., & Karray, S. 2022a. Applying Hybrid Machine Learning Algorithms To Assess Customer Risk-Adjusted Revenue In The Financial Industry. Electronic Commerce Research And Applications, 56, 101202. Amsterdam: Elsevier. Https://Doi.Org/10.1016/J.Elerap.2022.101202.

Machado, M. R., & Karray, S. 2022b. Assessing Credit Risk Of Commercial Customers Using Hybrid Machine Learning Algorithms. Expert Systems With Applications, 200, 116889. Amsterdam: Elsevier. Https://Doi.Org/10.1016/J.Eswa.2022.116889.

Ocktavia, S., & Atmojo, W. T. 2025. Analisis Segmentasi Pelanggan Pembiayaan Berdasarkan Demografi Untuk Memprediksi Tingkat Kredit Menggunakan Algoritma K-Means. Jurnal Informatika Teknologi Dan Sains, 7(1). Sumbawa: Universitas Teknologi Sumbawa. Https://Doi.Org/10.51401/Jinteks.V7i1.5582.

Rizkyanto, H., & Gaol, F. L. 2023. Customer Segmentation Of Personal Credit Using Recency, Frequency, Monetary And K-Means On Financial Industry. International Journal Of Advanced Computer Science And Applications, 14(4). London: The Science And Information Organization. Https://Doi.Org/10.14569/IJACSA.2023.0140417.

Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. 2022. Machine Learning-Driven Credit Risk: A Systematic Review. Neural Computing And Applications, 34, 14327–14339. London: Springer. Https://Doi.Org/10.1007/S00521-022-07472-2.

Wewengkang, F. S., & Wibowo, A. 2026. Analysis Of Online Loan Regional Clustering In Indonesia In 2024 Based On Outstanding And Default Rate Using K-Means Clustering. Al-Kharaj: Journal Of Islamic Economic And Business, 8(1). Palopo: IAIN Palopo. Https://Doi.Org/10.24256/Kharaj.V8i1.9342.

Yu, T., Huang, W., Tang, X., & Zheng, D. 2025. A Hybrid Unsupervised Machine Learning Model With Spectral Clustering And Semi-Supervised Support Vector Machine For

Credit Risk Assessment. Plos ONE, 20(1), E0316557. San Francisco: Public Library Of Science. Https://Doi.Org/10.1371/Journal.Pone.0316557.

Yuan, K., Chi, G., Zhou, Y., & Yin, H. 2022. A Novel Two-Stage Hybrid Default Prediction Model With K-Means Clustering And Support Vector Domain Description. Research In International Business And Finance, 59, 101536. Amsterdam: Elsevier. Https://Doi.Org/10.1016/J.Ribaf.2021.101536.