



UIN SUSKA RIAU

**SEGMENTASI UMKM KOTA PEKANBARU BERDASARKAN
PARTISIPASI DALAM E-KATALOG LOKAL
MENGGUNAKAN ALGORITMA CLUSTERING**

© Hak cipta milik UIN Suska Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer pada
Program Studi Sistem Informasi

Oleh:

RAHMA ALIYA

12250320351



State Islamic University of Sultan Syarif Kasim Riau

LEMBAR PERSETUJUAN

SEGMENTASI UMKM KOTA PEKANBARU BERDASARKAN PARTISIPASI DALAM E-KATALOG LOKAL MENGGUNAKAN ALGORITMA *CLUSTERING*

TUGAS AKHIR

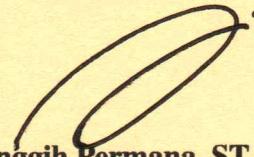
Oleh:

RAHMA ALIYA

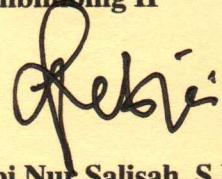
12250320351

Telah diperiksa dan disetujui sebagai laporan tugas akhir
di Pekanbaru, pada tanggal 20 Januari 2026

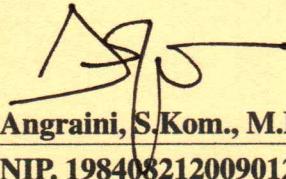
Pembimbing I


Inggih Permana, ST., M.Kom.
NIP. 198812102015031006

Pembimbing II


Febi Nur Salisah, S.Kom., M.Kom.
NIP. 19900222023212038

Ketua Program Studi


Angraini, S.Kom., M.Eng.,Ph.D.
NIP. 198408212009012008

LEMBAR PENGESAHAN

SEGMENTASI UMKM KOTA PEKANBARU BERDASARKAN PARTISIPASI DALAM E-KATALOG LOKAL MENGGUNAKAN ALGORITMA *CLUSTERING*

TUGAS AKHIR

Oleh:

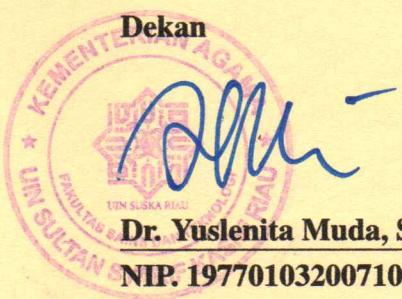
RAHMA ALIYA

12250320351

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 08 Januari 2026

Pekanbaru, 08 Januari 2026

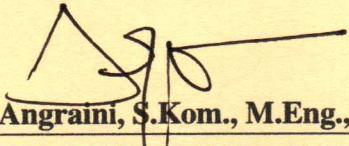
Mengesahkan,



Dekan

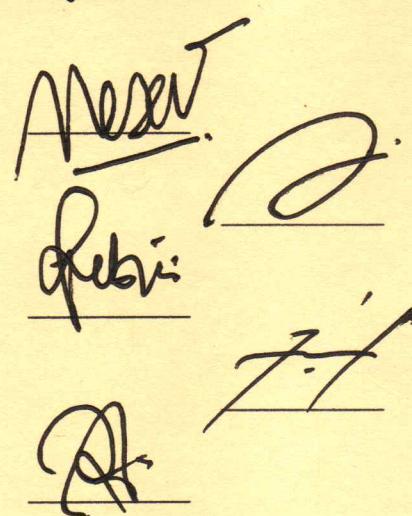
Dr. Yuslenita Muda, S.Si., M.Sc.
NIP. 197701032007102001

Ketua Program Studi


Angraini, S.Kom., M.Eng., Ph.D.
NIP. 198408212009012008

DEWAN PENGUJI:

Ketua : Nesdi Evrilyan Rozanda, S.Kom., M.Sc.



Sekretaris : Inggih Permana, ST., M.Kom.

Anggota 1 : Febi Nur Salisah, S.Kom., M.Kom.

Anggota 2 : Dr. Rice Novita, S.Kom., M.Kom.

Anggota 3 : M. Jazman, S.Kom., M.Infosys.

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Rahma Aliya

NIM : 12250320351

Tempat/Tgl. Lahir : Kampung Jawa, 02 November

Fakultas/Pascasarjana : Fakultas Sains dan Teknologi

Prodi : Sistem Informasi

Judul Skripsi

SEGMENTASI UMKM KOTA PEKANBARU BERDASARKAN PARTISIPASI DALAM E-KATALOG LOKAL MENGGUNAKAN ALGORITMA CLUSTERING

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan Skripsi dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu Skripsi saya ini, saya nyatakan bebas dari plagiat.
4. Apa bila dikemudian hari terbukti terdapat plagiat dalam penulisan Skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 22 Januari 2026

Yang membuat pernyataan



Rahma Aliya

NIM: 12250320351



LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum, dengan ketentuan bahwa hak cipta ada pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan atas izin penulis dan harus dilakukan mengikuti kaedah dan kebiasaan ilmiah serta menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin tertulis dari Dekan fakultas universitas. Perpustakaan dapat meminjamkan Tugas Akhir ini untuk anggotanya dengan mengisi nama, tanda peminjaman dan tanggal pinjam pada *form* peminjaman.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak meugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Pekanbaru, 20 Januari 2026

Yang membuat pernyataan,

RAHMA ALIYA

NIM. 12250320351

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengujikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSEMBAHAN

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillah wa syukurillah, sebagai bentuk rasa syukur peneliti atas segala nikmat yang telah didapatkan tanpa ada kekurangan sekalipun. Sholawat berserta salam tak lupa peneliti ucapkan kepada Nabi Muhammad Shalallahu 'Alaihi Wa Sallam dengan melaangkan Allahumma Shalli 'ala Sayyidina Muhammad wa 'ala Ali Sayyidina Muhammad. Semoga kita semua senantiasa mendapat syafaat-Nya di dunia maupun di akhirat. Amiin Ya Rabbal 'Alamiin.

Tugas Akhir ini peneliti persembahkan kepada Ayahanda Ranu dan Ibunda Maria Ulfa yang senantiasa memberikan kasih sayang, doa, serta dukungan moral maupun motivasi tanpa henti kepada peneliti, sehingga peneliti dapat menyelesaikan pendidikan Strata Satu (S-1) dengan baik. Berkat ketulusan, pengorbanan, dan kesabaran beliau, peneliti akhirnya mampu menyelesaikan Laporan Tugas Akhir ini.

Karya ini peneliti dedikasikan sebagai bentuk rasa syukur sekaligus tolak ukur keberhasilan yang peneliti persembahkan kepada keluarga besar tercinta. Peneliti juga mengucapkan terima kasih kepada kedua adik peneliti, Miftahul Jannah dan Zakiatul Ilma, yang senantiasa memberikan semangat, doa, dan dukungan selama proses penyusunan Tugas Akhir ini. Terima kasih kepada Bapak Inggih Permana, ST., M.Kom dan Ibu Febi Nur Salisah, S.Kom., M.Kom. yang telah sabar membimbing peneliti untuk menyelesaikan Tugas Akhir ini dan selalu menyempatkan waktu untuk mengajari serta membimbing peneliti dalam semua hal yang berkaitan dengan Tugas Akhir ini.

Semoga Allah Subhanahu Wa Ta'ala memberikan keberkahan serta rahmat-Nya dan senantiasa menjadikan kita hamba yang terus berdoa kepadaNya. Amiin Ya Rabbal 'Alamiin.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

UIN SUSKA RIAU

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengikuti kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

KATA PENGANTAR

Puji dan syukur peneliti panjatkan ke hadirat Allah Subhanahu wa Ta'ala atas rahmat dan karunia-Nya sehingga Laporan Tugas Akhir yang berjudul “Segmentasi UMKM Kota Pekanbaru Berdasarkan Partisipasi dalam E-Katalog Lokal Menggunakan Algoritma K-Means, DBSCAN, dan Hierarchical Clustering” dapat diselesaikan dengan baik. Laporan ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom.) pada Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau.

Peneliti menyadari bahwa penyusunan Laporan Tugas Akhir ini tidak terlepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, peneliti mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Ibu Prof. Dr. Hj. Leny Nofianti MS, SE, M.Si, Ak, CA., sebagai Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Ibu Dr. Yuslenita Muda, S.Si., M.Sc., sebagai Dekan Fakultas Sains dan Teknologi.
3. Ibu Angraini, S.Kom., M.Eng., Ph.D., sebagai Ketua Program Studi Sistem Informasi.
4. Ibu Dr. Rice Novita, S.Kom., M.Kom., sebagai Sekretaris Program Studi Sistem Informasi, sekaligus Dosen Penguji 1 yang telah banyak memberikan arahan, masukan, dan motivasi selama perkuliahan.
5. Bapak Inggh Permana, ST., M.Kom., selaku Dosen Pembimbing I yang telah memberikan bimbingan, arahan, serta masukan yang konstruktif, serta banyak memberikan motivasi dan dorongan kepada penulis sehingga penulis dapat menyelesaikan tugas akhir ini dengan tepat waktu.
6. Ibu Febi Nur Salisah, S.Kom., M.Kom., selaku Dosen Pembimbing 2 yang telah memberikan pendampingan, saran, dan motivasi kepada penulis dalam menyempurnakan tugas akhir ini.
7. Bapak M. Jazman, S.Kom., M.Infosys., sebagai Dosen penguji 2 tugas akhir ini, yang dengan penuh dedikasi dan ketelitian memberikan arahan, masukan, nasehat, serta motivasi dalam penyelesaian Tugas Akhir ini.
8. Bapak Nesdi Evrilyan Rozanda, S.Kom., M.Sc., sebagai Ketua Sidang yang telah dengan bijaksana memimpin sidang Tugas Akhir ini serta memberikan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengujikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

arahan maupun masukan yang sangat berarti demi perbaikan Tugas Akhir ini

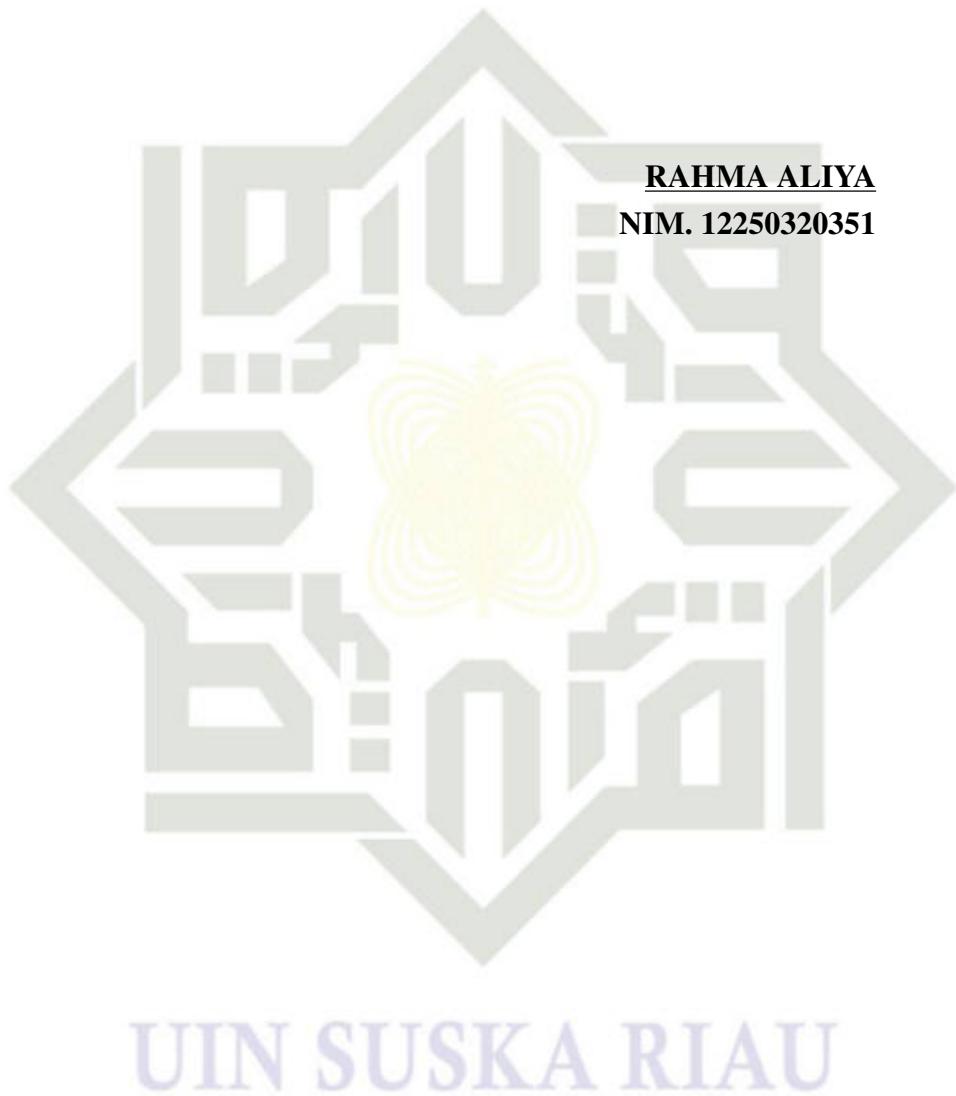
9. Ibu Fitriani Muttakin, S.Kom., M.Cs., sebagai Dosen Pembimbing Akademik peneliti yang telah banyak memberikan arahan, masukan, dan motivasi dalam penyelesaian Tugas Akhir ini
 10. Kedua orang tua tercinta, Ayah Ranu dan Ibu Maria Ulfa, yang telah memberikan dukungan moral maupun material, semangat, pengorbanan, serta doa yang tiada henti sejak peneliti kecil hingga mampu menyelesaikan pendidikan Strata Satu (S-1) dan Laporan Tugas Akhir ini.
 11. Kedua adik tercinta, Miftahul Jannah dan Zakiatul Ilma, yang senantiasa menjadi sumber motivasi dan penyemangat bagi peneliti dalam menyelesaikan Laporan Tugas Akhir ini.
 12. Bapak dan Ibu dosen serta seluruh pegawai dan pegawai/i pada pihak Kemahasiswaan UIN Sultan Syarif Kasim Riau yang telah memberikan pelayanan, bimbingan, serta akses kepada peneliti selama masa perkuliahan hingga penyelesaian Laporan Tugas Akhir.
 13. Abang Alvin Afwan dan seluruh Pegawai Dinas Koperasi Usaha Kecil dan Menengah Kota Pekanbaru, yang telah banyak membantu peneliti selama proses penelitian, khususnya dalam memberikan arahan, masukan, dan dukungan sehingga penelitian ini dapat berjalan dengan baik.
 14. Sahabat peneliti, Citra Ardila, Nurwahid Hidayat Syam, dan Fitri Rahmadian, yang senantiasa memberikan dukungan moral, semangat, dan motivasi selama proses penyusunan Tugas Akhir ini.
 15. Teman-teman dari Grup Ciwi Pikmi dan Grup Bismillah yang tidak dapat disebutkan satu per satu, terima kasih atas dukungan, kebersamaan, serta semangat yang telah diberikan kepada peneliti.
 16. Teman-teman Program Studi Sistem Informasi Kelas C yang telah memberikan kebersamaan, bantuan, dan motivasi selama masa perkuliahan dan proses penyusunan Laporan Tugas Akhir.
 17. Teman-teman Kuliah Kerja Nyata (KKN) Desa Pesaguan yang telah memberikan pengalaman, kerja sama, dan dukungan selama kegiatan KKN berlangsung.
 18. Seluruh pihak yang tidak dapat peneliti sebutkan satu per satu yang telah banyak membantu peneliti dalam proses penyelesaian Tugas Akhir ini.
- Peneliti menyadari bahwa Laporan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang bersifat membangun sangat di-

harapkan. Apabila terdapat pertanyaan terkait Tugas Akhir ini, peneliti dapat dihubungi melalui surel 12250320351@students.uin-suska.ac.id. atau rahmaalia813@gmail.com. Akhir kata, semoga Tugas Akhir ini dapat memberikan manfaat bagi semua pihak yang membutuhkan serta menjadi referensi bagi pengembangan ilmu pengetahuan di bidang terkait.

Pekanbaru, 20 Januari 2026

Penulis,

RAHMA ALIYA
NIM. 12250320351



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



UIN SUSKA RIAU

SEGMENTASI UMKM KOTA PEKANBARU BERDASARKAN PARTISIPASI DALAM E-KATALOG LOKAL MENGGUNAKAN ALGORITMA CLUSTERING

RAHMA ALIYA
NIM: 12250320351

Tanggal Sidang: 08 Januari 2026

Periode Wisuda:

Program Studi Sistem Informasi
Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. Soebrantas, No. 155, Pekanbaru

ABSTRAK

Penelitian ini bertujuan untuk melakukan segmentasi UMKM Kota Pekanbaru berdasarkan tingkat partisipasi dalam e-Katalog Lokal menggunakan algoritma K-Means, DBSCAN, dan Hierarchical Clustering. Data penelitian terdiri dari 10.560 UMKM dengan variabel omzet, jumlah tenaga kerja, lama berdiri usaha, dan partisipasi e-Katalog Lokal. Hasil analisis menunjukkan bahwa K-Means menghasilkan tiga klaster UMKM dengan karakteristik partisipasi e-Katalog yang berbeda. Algoritma DBSCAN membentuk dua klaster tanpa menghasilkan data noise dengan nilai Silhouette Score sebesar 0,666, sedangkan Hierarchical Clustering menghasilkan struktur klaster bertingkat dengan nilai Silhouette Score sebesar 0,693. Meskipun demikian, K-Means dinilai paling stabil dan mudah diinterpretasikan untuk segmentasi berbasis kebijakan publik.

Kata Kunci: *UMKM, e-Katalog Lokal, K-Means Clustering, DBSCAN, Hierarchical Clustering, segmentasi data*

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak mengujikan kepentingan yang wajar UIN Suska Riau.

© Hak cipta milik UIN Suska Riau



UIN SUSKA RIAU

SEGMENTATION MSMES IN PEKANBARU CITY BASED ON PARTICIPATION IN LOCAL E-CATALOGS USING K-MEANS DBSCAN AND HIERARCHICAL CLUSTERING ALGORITHMS

RAHMA ALIYA
NIM: 12250320351

Date of Final Exam: January 08th 2026
Graduation Period:

Department of Information System
Faculty of Science and Technology
State Islamic University of Sultan Syarif Kasim Riau
Soebrantas Street, No. 155, Pekanbaru

ABSTRACT

This study aims to segment MSMEs in Pekanbaru City based on their level of participation in the Local e-Catalog using the K-Means, DBSCAN, and Hierarchical Clustering algorithms. The research data consists of 10,560 MSMEs with variables of turnover, number of employees, length of business establishment, and participation in the Local e-Catalog. The analysis results show that K-Means produces three clusters of MSMEs with different characteristics of e-Catalog participation. The DBSCAN algorithm formed two clusters without generating noise data with a Silhouette Score of 0.666, while Hierarchical Clustering produced a hierarchical cluster structure with a Silhouette Score of 0.693. However, K-Means was considered the most stable and easy to interpret for public policy-based segmentation.

Keywords: Data Segmentation, DBSCAN, Hierarchical Clustering, K-Means Clustering, Local e-Catalog, MSMEs,

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau

UIN SUSKA RIAU

DAFTAR ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMAHAN	vi
KATA PENGANTAR	vii
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI	xii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvi
DAFTAR SINGKATAN	xvii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	7
1.3 Batasan Masalah	8
1.4 Tujuan	8
1.5 Manfaat	9
1.6 Sistematika Penulisan	9
2 LANDASAN TEORI	11
2.1 UMKM	11
2.2 Digitalisasi UMKM dan E-Katalog Lokal	15
2.3 Data Mining	18
2.4 Clustering	21
2.5 Algoritma K-Means Clustering	23
2.5.1 Menentukan jumlah cluster yang mau di bentuk	24
2.5.2 Menentukan centeroid awal secara acak	24
2.5.3 Menghitung jarak setiap data ke masing-masing centroid dari masing-masing Clustering	25
2.5.4 Fungsi Tujuan	25
2.5.5 Memperbarui Centroid	26
2.6 Agoritma DBSCAN	27

2.6.1	Konsep Density-Based Clustering	27
2.6.2	Parameter Eps Dan Minpts	27
2.6.3	Tipe-Tipe Titik: Core Border, Noise	28
2.6.4	Mekanisme Pembentukan Cluster	28
2.7	Hierarchical Clustering	29
2.7.1	Konsep Agglomerative dan Divisive	30
2.7.2	Linkage Methods	30
2.7.3	Dendrogram	31
2.7.4	Kelebihan Dan Kelemahan Kelebihan	32
2.8	Evaluasi Kualitas Clustering	33
2.8.1	Silhouette Coefficient	33
2.9	Orange	36
2.10	Penelitian Terdahulu	38
3	METODOLOGI PENELITIAN	47
3.1	Alur Penelitian	47
3.1.1	3.1.1. Tahap Perencanaan	47
3.1.2	Tahap Pengumpulan Data	48
3.1.3	Tahap Segmentasi	49
3.1.3.1	Pra Pemrosesan Data	50
3.1.3.2	Pemodelan Dengan Algoritma K-Means Clustering	51
3.1.3.3	Pemodelan Dengan Algoritma Dbscan	52
3.1.3.4	Pemodelan Dengan Algoritma Hierarchical Clustering	53
3.1.4	Evaluasi	55
3.2	Jenis Penelitian	55
3.3	Alat dan Perangkat Lunak	56
3.4	Lokasi dan Waktu Penelitian	57
4	HASIL DAN PEMBAHASAN	58
4.1	Gambaran Umum Data	58
4.2	Hasil Tahap Perencanaan	59
4.2.1	Hasil Penentuan Objek Penelitian	59
4.2.2	Hasil Identifikasi Masalah	59
4.2.3	Hasil Studi Literatur	60
4.3	Hasil Pengumpulan Data	64
4.3.1	Data Dinas Koperasi Dan UMKM Kota Pekanbaru	64
4.3.2	LKPP (Data Partisipasi E-Katalog Lokal)	65
4.3.3	Data Dari Website E-Katalog Lokal (Inaproc.id)	66
4.4	Hasil Integrasi Data	67
4.5	Hasil Clustering	70
4.5.1	Hasil Processing Data	70
4.5.1.1	Hasil Imputasi Nilai Hilang (Impute Missing Values)	71
4.5.1.2	Hasil Normalisasi Variabel Numerik (Normalize Features)	71

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.5.1.3	Hasil Transformasi Variabel Kategorik (Continuize Discrete Variables)	71
4.5.2	Hasil Pengclusteran Algoritma	75
4.5.2.1	Hasil Algoritma K-Means Clustering	75
4.5.2.2	Hasil Algoritma DBSCAN	80
4.5.2.3	Hasil Algoritma Hierachial Clustering	86
4.6	Perbandingan Hasil <i>Cluster</i>	91
PENUTUP		95
5.1	Kesimpulan	95
5.2	Saran	95

DAFTAR PUSTAKA

LAMPIRAN A	Surat Izin Penelitian dan pengambilan data	A - 2
LAMPIRAN B	Surat Jawaban Izin Penelitian dan Pengambilan Data	B - 1
LAMPIRAN C	Dokumentasi Pengambilan Data di LPSE	C - 1
LAMPIRAN D	Dokumentasi Pengambilan Data di Dinas Koperasi	D - 1

UIN SUSKA RIAU

DAFTAR GAMBAR

2.1	Kegiatan bazar UMK	11
2.2	Grafik UMKM di Indonesia	13
2.3	Logo E-Katalog	16
2.4	Grafik jumlah UMKM di Pekanbaru tahun 2023	17
2.5	Komponen sistem informasi	19
2.6	Orange Data Mining	36
3.1	Alur Penelitian	47
3.2	Tahap <i>Clustering</i>	50
4.1	Visualisasi silhouette score K-means	79
4.2	Visualisasi silhouette score K-means	80
4.3	Visualisasi silhouette score DBSCAN	85
4.4	Dendogram dengan Max Depth 5	88
4.5	Visualisasi silhouette score HC	91

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR TABEL

2.1	Klasifikasi Usaha Berdasarkan Undang-Undang	12
2.2	Interpretasi Nilai Silhouette	34
2.3	Penelitian Terdahulu Terkait UMKM, E-Katalog, dan Clustering	38
4.1	Ringkasan Penelitian Terkait	61
4.2	Ringkasan Jenis Data dari Dinas Koperasi dan UKM Kota Pekanbaru	64
4.3	Data UMKM Dinas Koperasi dan UKM Kota Pekanbaru	65
4.4	Data dari LKPP Pekanbaru	66
4.5	Data dari LKPP Pekanbaru	66
4.6	Data UMKM pada e-Katalog Lokal (inaproc.id)	67
4.7	Primary Key	68
4.8	Contoh Dataset UMKM	68
4.9	Variabel Penelitian dan Sumber Data	69
4.10	Data UMKM Setelah Normalisasi dan Encoding	72
4.11	Statistik Deskriptif Variabel Penelitian	73
4.12	Hasil Algoritma K-Means	76
4.13	Hasil Cluster K-Means	77
4.14	Hasil Evaluasi K-Means	78
4.15	Hasil Evaluasi Parameter DBSCAN	81
4.16	Hasil Clustering UMKM Menggunakan DBSCAN	82
4.17	Hasil Cluster DBSCAN	83
4.18	Hasil Pemrosesan HC	86
4.19	Hasil Clustering HC	89
4.20	Evaluasi Shiloet score Hc	90
4.21	Perbandingan Hasil Clustering	92
4.22	Perbandingan Evaluasi	92
4.23	Hasil Cluster Terbaik	93

DAFTAR SINGKATAN

BI	:	Bankindonesia
BLP	:	Bela Pengadaan Lokal
BPS	:	Badan Pusat Statistik
CHI	:	Calinski–Harabasz Index
COVID-19	:	Coronavirus Disease 2019
CSV	:	Comma-Separated Values
DBI	:	Davies–Bouldin Index
DBSCAN	:	Density-Based Spatial Clustering of Applications with Noise
HC	:	Hierarchical Clustering
KADIN	:	Kamar Dagang dan Industri
LKPP	:	Lembaga Kebijakan Pengadaan Barang/Jasa Pemerintah
MSMEs	:	Micro, Small, and Medium Enterprises
PCA	:	Principal Component Analysis
PDB	:	Produk Domestik Bruto
SLR	:	Systematic Literature Review
SME	:	Small and Medium Enterprises
UKM	:	Usaha Kecil dan Menengah
UMKK	:	Usaha Mikro dan Kecil
UMKM	:	Usaha Mikro, Kecil, dan Menengah

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengikuti kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan bentuk badan usaha perseorangan yang memiliki peranan strategis dalam perekonomian nasional. Meskipun berskala relatif kecil, UMKM mampu menghasilkan produk-produk yang memiliki daya saing dan diminati oleh pasar internasional, sehingga secara tidak langsung berkontribusi terhadap peningkatan nilai ekspor serta memperkuat neraca perdagangan Indonesia (Syarifah, Purnamasari, dan Purnomo, 2021). Selain itu, UMKM menjadi salah satu pilar utama dalam struktur ekonomi nasional karena keberadaannya yang tersebar di seluruh wilayah Indonesia, yang mencerminkan aktivitas ekonomi masyarakat pada tingkat akar rumput. Melalui kontribusinya dalam penyerapan tenaga kerja, peningkatan pendapatan masyarakat, serta pemerataan pembangunan ekonomi, UMKM berperan penting dalam menjaga stabilitas dan mendorong pertumbuhan ekonomi nasional secara berkelanjutan (Jelita, Buation, dan Simanjuntak, 2023).

Usaha mikro, kecil, dan menengah (UMKM) memegang peranan penting dalam menyerap lebih dari 97% tenaga kerja di Indonesia yang berdampak pada 57% produk domestik bruto (PDB) (Badan Pusat Statistik, 2021). Keberadaan UMKM dalam struktur perekonomian Indonesia menempati posisi yang sangat penting(Alamanda, Kusmiati, Shiddieq, dan Roji, 2023). Dalam konteks pengembangan UMKM, sudah banyak yang mulai mengadopsi penggunaan teknologi informasi yang dapat meningkatkan efisiensi dan efektivitas proses yang ada. Dalam skala yang lebih luas, penggunaan teknologi digital dapat meningkatkan kinerja bisnis secara keseluruhan.Serta mampu meningkatkan kemampuan UMKM dalam memperoleh atau menyampaikan informasi untuk pasar.

Salah satu bentuk upaya pemerintah Indonesia dalam memanfaatkan teknologi sebagai media pengembangan UMKM adalah pemanfaatan katalog elektronik dalam pemanfaatan barang dan jasa pemerintah. Katalog elektronik sebagai bagian dari proses perdagangan secara elektronik diyakini mampu membuka pasar baru (Hamadt al., 2018). Katalog elektronik merupakan salah satu bagian dari e-marketplace dalam proses pengadaan barang dan jasa, yaitu pasar elektronik yang disediakan untuk memenuhi kebutuhan barang dan jasa pemerintah. (LKPP, 2018). Dengan cakupan pasar dan karakteristik yang beragam, katalog elektronik dibedakan menjadi beberapa jenis, yaitu katalog elektronik nasional, sektoral, dan

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengujikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

lokal. Katalog elektronik nasional disusun dan dikelola oleh Lembaga Kebijakan Pengadaan Barang/Jasa Pemerintah, katalog elektronik sektoral yang memuat informasi barang/jasa umum dan inovasi yang dikelola oleh kementerian/lembaga, dan katalog elektronik daerah atau lokal yang dikelola oleh pemerintah daerah.

Dalam konteks kebijakan nasional, Peraturan Presiden Nomor 12 Tahun 2021 tentang Perubahan atas Perpres Nomor 16 Tahun 2018 mengamanatkan bahwa kementerian/lembaga dan pemerintah daerah wajib mengalokasikan minimal 40 persen anggaran pengadaan barang dan jasa untuk produk Usaha Mikro, Kecil, dan Koperasi (UMKK) (Sekretariat Kabinet RI, 2021). Kebijakan ini menaikkan batasan paket pengadaan untuk usaha mikro dan kecil menjadi Rp15 miliar, enam kali lipat dari nilai sebelumnya sebesar Rp2,5 miliar, sebagai upaya memberikan kesempatan lebih luas bagi UMKM untuk berpartisipasi dalam pengadaan pemerintah (BLP Malang, 2021). Namun, dari sekitar 66 juta pelaku UMKM di Indonesia berdasarkan data BPS 2023, realisasi produk UMKM yang tayang di e-katalog nasional baru mencapai 7,6 juta produk (LKPP, 2023), menunjukkan kesenjangan masif antara potensi dan adopsi aktual. Kondisi ini mengindikasikan urgensi penelitian yang dapat mengidentifikasi hambatan spesifik dan merumuskan strategi intervensi berbasis data, khususnya pada konteks lokal seperti Kota Pekanbaru.

E-katalog lokal atau daerah hadir sebagai salah satu solusi strategis untuk meminimalisir praktik penipuan dalam proses pengadaan barang dan jasa. Namun, meskipun menawarkan transparansi dan efisiensi, tingkat adopsi e-katalog di kalangan pelaku usaha masih tergolong rendah. Padahal, platform ini memiliki berbagai manfaat, terutama bagi pemangku kebijakan dan pihak yang memiliki kewenangan dalam proses pengadaan, karena mampu mempercepat proses, meningkatkan akurasi, serta mengurangi potensi penyimpangan (Maahira, 2023).

Namun, keterlibatan UMKM dalam platform digital pemerintah seperti e-katalog masih rendah. Padahal, sistem ini penting untuk meningkatkan transparansi dan efisiensi pengadaan barang dan jasa publik. Minimnya partisipasi menimbulkan kesenjangan digital dan membatasi peluang pasar UMKM lokal. Meskipun strategis, partisipasi UMKM dalam sistem digital pemerintah seperti e-katalog lokal masih sangat terbatas. Platform ini sejatinya dirancang untuk mendukung pengadaan barang dan jasa publik secara transparan dan efisien, namun banyak UMKM yang belum terintegrasi karena kendala literasi digital dan kapasitas sumber daya (Fahlevi, Putri, dan Syahrin, 2023). Di Kota Pekanbaru, kondisi ini menciptakan kesenjangan digital yang membatasi potensi pemasaran UMKM melalui jalur pengadaan pemerintah (Budiono, Nugroho, dan Santoso, 2024). Di Kota Pekan-

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengujikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

baru, kondisi ini menciptakan kesenjangan digital yang signifikan. Berdasarkan data Dinas Koperasi dan UKM Kota Pekanbaru, jumlah UMKM tercatat mencapai 26.684unit usaha pada tahun 2024, meningkat dari 25.074unit pada tahun 2023, dengan dominasi sektor kuliner dan ekonomi kreatif (Sarbaini, 2024). Namun, tingkat partisipasi UMKM dalam e-katalog lokal Pekanbaru masih sangat rendah, dengan kesenjangan substansial antara jumlah UMKM terverifikasi dan UMKM yang aktif bertransaksi melalui platform e-katalog (Ahmad, 2024) Kondisi ini membatasi potensi pemasaran UMKM melalui jalur pengadaan pemerintah dan menghambat realisasi target Peraturan Presiden No. 12 Tahun 2021 yang mewajibkan alokasi minimal 40% anggaran pengadaan barang dan jasa pemerintah untuk UMKM.

Oleh karena itu, segmentasi UMKM berdasarkan karakteristik usaha merupakan pendekatan yang strategis dalam mendukung proses pengambilan kebijakan yang lebih tepat dan terarah. Selama ini, intervensi pemerintah terhadap UMKM, seperti program pelatihan atau bantuan pembiayaan, cenderung dilakukan secara umum tanpa mempertimbangkan kesiapan digital, kapasitas usaha, atau tingkat partisipasi dalam platform e-katalog. Hal ini menimbulkan ketidaktepatan sasaran, rendahnya efektivitas program, dan pemborosan sumber daya. Melalui segmentasi berbasis data, pemerintah dapat memperoleh gambaran yang lebih jelas mengenai profil, kebutuhan, serta potensi dari masing-masing kelompok usaha, sehingga setiap klaster dapat diberikan intervensi yang sesuai dengan kondisi aktualnya. Dengan demikian, segmentasi berbasis data menjadi landasan penting dalam mewujudkan kebijakan yang responsif terhadap keragaman dan dinamika sektor UMKM (Iin, Fadila, Rinaldi, dan Fathurrohman, 2024).

Untuk memahami karakteristik UMKM yang belum berpartisipasi dalam e-katalog lokal, diperlukan pendekatan analitis berbasis data. Clustering merupakan pendekatan analitik yang umum digunakan untuk segmentasi data. Salah satu algoritma yang banyak digunakan adalah K-Means, namun pendekatan lain seperti Hierarchical Clustering dan DBSCAN juga relevan untuk memberikan perspektif komparatif dalam melihat pola UMKM untuk segmentasi berdasarkan parameter seperti omzet, jumlah tenaga kerja, jenis produk, dan status digitalisasi (Mawarni, Testiana, dan Dalafranka, 2023). Dengan pendekatan ini, pemerintah dapat merumuskan intervensi yang sesuai untuk tiap kelompok UMKM sesuai kesiapan digital dan kapasitas bisnisnya (Alzami dkk., 2022).

Metode K-Means Clustering merupakan salah satu algoritma yang banyak digunakan dalam segmentasi data karena prosesnya yang sederhana dan efisien, terutama dalam menangani data berjumlah besar. algoritma yang banyak digunakan

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengikuti kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dalam segmentasi data karena prosesnya yang sederhana dan efisien, terutama dalam menangani data berjumlah besar. Dalam konteks pengelompokan UMKM, algoritma ini diterapkan untuk mengklasifikasikan unit usaha berdasarkan kriteria seperti omzet, jumlah tenaga kerja, dan jenis usaha. Hasil segmentasi yang diperoleh mampu menunjukkan perbedaan karakteristik antar kelompok, misalnya adanya klaster UMKM dengan omzet tinggi namun jumlah tenaga kerja yang sedikit, serta sebaliknya. Informasi ini dapat dimanfaatkan untuk merancang strategi pembinaan dan pengembangan kebijakan yang lebih tepat sasaran, sesuai dengan kebutuhan dari masing-masing segmen usaha(Azzam, Purnamasari, dan Ali, 2024).

Selain K-Means, algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) menawarkan pendekatan berbasis kepadatan yang mampu menangani outlier dan noise dalam data UMKM. Studi oleh (A. Ramadhan, Achmad, Zulkarnain, dan Aritsugi, 2025)menerapkan DBSCAN untuk clustering MSME di Sidoarjo berdasarkan modal, turnover, dan tenaga kerja, dengan hasil menunjukkan kemampuan superior dalam menangani variasi kepadatan dan mengidentifikasi outlier. Wikarta / Defiyanti (2025) dalam evaluasi komparatif di Rembang menemukan bahwa DBSCAN menghasilkan kualitas clustering terbaik dengan Silhouette Score sebesar 0.5496, lebih unggul dibandingkan K-Means terutama dalam menangani noise dan cluster berbentuk tidak beraturan. (Sutramiani dkk., 2024)juga membuktikan keunggulan DBSCAN dengan preprocessing PCA untuk identifikasi cluster industri mikro/kecil, menangani sparsity dan multikolinearitas data spasial dengan efektif.

Sementara itu, Hierarchical Clustering (agglomerative maupun divisive) menyediakan pendekatan alternatif yang menghasilkan visualisasi hierarkis berupa dendrogram untuk memahami relasi struktural antar cluster(Wijaya, Budajati, dan Wicaksono, 2025) mengimplementasikan agglomerative hierarchical clustering dengan complete-linkage untuk mengelompokkan usaha mikro kecil di Jawa Timur berdasarkan data BPS, menghasilkan identifikasi kelompok usaha dengan karakteristik produktivitas berbeda.(Selayanti, Putri, dan Fahrudin, 2025) menerapkan Agglomerative Hierarchical Clustering (AHC) untuk segmentasi Sentra Wisata Kuliner UMKM di Surabaya menggunakan cophenetic correlation dan validasi Silhouette Coefficient, berhasil membentuk 3 cluster yang merepresentasikan tingkat perkembangan usaha yang berbeda.(Wikarta dan Defiyanti, 2025)mencatat bahwa meskipun hierarchical clustering menunjukkan kualitas menengah, metode ini memberikan keunggulan dalam visualisasi relasi hierarkis antar cluster dan kemu-

dahan interpretasi hasil, menjadikannya alternatif yang sesuai untuk dataset dengan struktur hierarkis yang jelas.

Penggunaan pendekatan komparatif dengan tiga algoritma clustering (K-Means, DBSCAN, Hierarchical) dalam penelitian ini memiliki justifikasi metodologis yang kuat. Studi komparatif oleh (Wardani, Ariyanto, Umroh, dan Rolliwati, 2023) pada data marketplace elektronik menunjukkan bahwa setiap algoritma memiliki karakteristik dan keunggulan berbeda: K-Means efisien untuk data terstruktur dengan cluster spherical dan dataset besar, DBSCAN unggul dalam menangani outlier dan cluster berbentuk arbitrer tanpa asumsi jumlah cluster awal, sementara Hierarchical Clustering memberikan interpretabilitas hierarkis yang mendalam meskipun dengan kompleksitas komputasi lebih tinggi. (John, Shobayo, dan Ogunleye, 2023) dalam eksplorasi algoritma clustering untuk customer profiling menegaskan bahwa seleksi algoritma optimal harus mempertimbangkan karakteristik intrinsik dataset, termasuk distribusi data, kehadiran outlier, bentuk geometri cluster, dan tujuan analitik spesifik. Dengan membandingkan ketiga algoritma tersebut, penelitian ini dapat mengidentifikasi metode yang paling optimal untuk karakteristik data UMKM Pekanbaru, sekaligus memberikan rekomendasi metodologis bagi penelitian selanjutnya dalam konteks segmentasi UMKM berbasis e-katalog.

Penelitian terdahulu telah banyak menunjukkan efektivitas algoritma K-Means Clustering dalam melakukan segmentasi terhadap UMKM dan berbagai entitas bisnis lainnya. Studi oleh Salsabiela et al. (2024) mengelompokkan produk toko UMKM berdasarkan kebutuhan restock dengan nilai Davies-Bouldin Index (DBI) sebesar 0,436 yang menandakan segmentasi yang cukup baik. Penelitian (Mawarni dkk., 2023) juga menunjukkan bahwa metode K-Means mampu mengelompokkan pelanggan berdasarkan nilai pembelian, menghasilkan kelompok pelanggan prioritas yang bermanfaat bagi strategi pemasaran. Temuan serupa disampaikan oleh (Afrizal, Saputra, dan Satria, 2025) yang menerapkan K-Means untuk segmentasi pasar UMKM dengan hasil Silhouette Score sebesar 0,72 dan DBI sebesar 0,45, menandakan model yang stabil dan representatif. Sementara itu, Arrahmi (2025) dan (Tertiaavini, 2024) mengaplikasikan K-Means pada konteks UMKM daerah dengan menggunakan variabel seperti sektor industri, jumlah tenaga kerja, dan omzet. Penelitian mereka berhasil membentuk klaster usaha mikro, kecil, dan menengah dengan karakteristik ekonomi berbeda, memberikan dasar bagi penyusunan kebijakan pembinaan UMKM yang lebih tepat sasaran.

Di sisi lain, sejumlah penelitian menyoroti pentingnya digitalisasi melalui e-katalog pemerintah sebagai sarana perluasan pasar UMKM. (Ahmad, 2024) meneliti

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengujikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

implementasi e-katalog di Kota Pekanbaru dan menemukan bahwa sistem ini meningkatkan transparansi serta efisiensi proses pengadaan barang/jasa, namun tingkat partisipasi UMKM masih rendah karena keterbatasan modal, literasi digital, dan akses informasi. Penelitian oleh (Sinta Puspita Sari dkk., 2024) meninjau regulasi dan prosedur menjadi penyedia e-katalog di bawah LKPP, menunjukkan bahwa kompleksitas administratif masih menjadi hambatan utama. Penelitian internasional oleh (Khan, Bashir, Batool, Raza, dan Bashir, 2024) memperkenalkan metode Differentiation-based Initialization untuk memperbaiki kelemahan inisialisasi centroid pada K-Means, yang menunjukkan upaya global dalam meningkatkan akurasi dan stabilitas metode ini.

Berdasarkan telaah sistematis terhadap 34 penelitian terdahulu, dapat disimpulkan bahwa kajian mengenai digitalisasi UMKM dan penerapan clustering telah memberikan kontribusi signifikan terhadap pengembangan strategi usaha kecil Nama-nama, sebagian besar penelitian masih berjalan secara terpisah: penelitian tentang e-katalog cenderung bersifat deskriptif dan berfokus pada kebijakan, sedangkan penelitian tentang clustering umumnya menyoroti karakteristik usaha tanpa mempertimbangkan aspek digitalisasi.

Dari sinilah muncul celah penelitian (research gap) yang ingin diisi oleh studi ini, yaitu belum adanya penelitian yang secara langsung menghubungkan tingkat partisipasi UMKM dalam e-katalog lokal dengan hasil segmentasi berbasis algoritma clustering multi-metode (K-Means, DBSCAN, Hierarchical). Penelitian ini berupaya mengisi kekosongan tersebut dengan menganalisis data UMKM di Kota Pekanbaru menggunakan pendekatan data mining untuk mengelompokkan UMKM berdasarkan omzet, jumlah tenaga kerja, lama berdiri, dan tingkat partisipasi dalam e-katalog lokal. Proses evaluasi dilakukan dengan Silhouette Coefficient dan Davies-Bouldin Index untuk memastikan kualitas dan robustness klaster yang terbentuk dari ketiga algoritma tersebut.

Dengan pendekatan tersebut, penelitian ini diharapkan dapat menghasilkan peta segmentasi UMKM digital Pekanbaru yang akurat dan aplikatif, serta memberikan rekomendasi strategis bagi pemerintah daerah dalam meningkatkan literasi digital dan partisipasi UMKM di platform e-katalog. Hasil penelitian ini juga diharapkan menjadi kontribusi nyata dalam memperkuat literatur mengenai integrasi e-procurement dan data-driven clustering dalam konteks ekonomi digital di Indonesia.

Penelitian ini menggunakan data sekunder UMKM Kota Pekanbaru yang diperoleh dari Dinas Koperasi dan UKM Kota Pekanbaru, mencakup variabel kuan-

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
 - b. Pengutipan tidak mengikuti kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

titatif meliputi omzet tahunan, jumlah tenaga kerja, lama berdiri usaha (dalam tahun), serta variabel kategorikal meliputi jenis usaha, sektor industri, status legalitas, dan status partisipasi dalam e-katalog lokal. Data tersebut akan diproses melalui tahapan preprocessing (normalisasi dan transformasi), dianalisis menggunakan tiga algoritma clustering (K-Means dengan inisialisasi K-Means++, DBSCAN dengan parameter ϵ dan MinPts yang dioptimasi, dan Hierarchical Clustering dengan linkage method yang sesuai), kemudian divalidasi menggunakan Silhouette Coefficient dan Davies-Bouldin Index untuk memastikan robustness dan kualitas segmentasi yang dihasilkan(Y. Wang, 2023).

Dengan desain penelitian berbasis multi-algoritma dan validasi ganda tersebut, Penelitian ini berkontribusi dalam pengembangan strategi berbasis bukti (evidence-based strategy) dengan memetakan UMKM di Pekanbaru berdasarkan keterlibatan dalam platform e-katalog. Dimensi ini belum banyak dieksplorasi di tingkat lokal di Indonesia, khususnya dengan pendekatan berbasis data mining (Pambudi, Witanti, 2021). Secara teoretis, penelitian ini mengisi kesenjangan literatur (research gap) dengan mengintegrasikan analisis partisipasi e-katalog dan clustering multi-algoritma pada konteks UMKM local pendekatan integrasi yang belum pernah dilakukan sebelumnya di Kota Pekanbaru maupun dalam literatur nasional yang lebih luas. Penelitian ini juga memberikan kontribusi metodologis melalui analisis komparatif performa tiga algoritma clustering (K-Means, DBSCAN, Hierarchical) pada dataset UMKM lokal, yang dapat menjadi referensi bagi penelitian sejenis di wilayah lain. Secara praktis, hasil segmentasi UMKM akan menjadi instrumen evidence-based policy bagi Dinas Koperasi dan UMKM Kota Pekanbaru dalam merancang program pembinaan yang lebih tepat sasaran dan efisien, terutama untuk meningkatkan literasi digital, kapasitas usaha, dan partisipasi UMKM dalam e-katalog lokal sesuai dengan profil klaster yang teridentifikasi. Dengan demikian, penelitian ini diharapkan menjadi kontribusi nyata dalam memperkuat ekosistem digitalisasi UMKM dan mendukung transformasi ekonomi digital yang inklusif dan berkelanjutan di Indonesia.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah. Bagaimana penerapan dan perbandingan algoritma K-Means, Hierarchical Clustering, dan DBSCAN dalam melakukan segmentasi UMKM di Kota Pekanbaru berdasarkan karakteristik usaha dan tingkat partisipasinya dalam e-katalog lokal, serta algoritma mana yang menghasilkan kualitas klaster paling optimal

berdasarkan metrik evaluasi yang digunakan?

1.3 Batasan Masalah

Agar penelitian ini lebih fokus dan terarah, maka ditetapkan beberapa batasan masalah sebagai berikut:

1. Penelitian ini difokuskan pada data Usaha Mikro, Kecil, dan Menengah (UMKM) yang beroperasi di wilayah administrasi Kota Pekanbaru, yang merupakan salah satu pusat pertumbuhan ekonomi di Provinsi Riau.
2. Segmentasi UMKM dilakukan berdasarkan variabel-variabel utama, yaitu tingkat partisipasi dalam e-Katalog lokal, omzet, jumlah tenaga kerja, lama berdiri usaha, kategori usaha.
3. Algoritma yang digunakan dalam penelitian ini menggunakan tiga algoritma clustering, yaitu K-Means, Hierarchical Clustering, dan DBSCAN, untuk memperoleh hasil segmentasi yang komprehensif
4. Penelitian ini mencoba beberapa variasi jumlah klaster (nilai k) pada algoritma K-Means, yaitu $k = 2$ hingga $k = 5$, untuk menemukan nilai k optimal menggunakan metode evaluasi Silhouette Score.
5. Penelitian ini tidak mencakup pengembangan atau pembuatan sistem aplikasi; fokus penelitian terbatas pada analisis data dan interpretasi hasil segmentasi.
6. Data yang digunakan dalam penelitian ini merupakan data sekunder tahun 2024, yang diperoleh dari Dinas Koperasi, Usaha Kecil dan Menengah (UKM) Kota Pekanbaru serta Layanan Pengadaan Secara Elektronik (LPSE), dengan total 10.560 data UMKM terverifikasi.

1.4 Tujuan

Penelitian ini bertujuan untuk:

1. Menerapkan dan membandingkan algoritma K-Means, Hierarchical Clustering, dan DBSCAN dalam melakukan segmentasi UMKM di Kota Pekanbaru berdasarkan karakteristik usaha serta tingkat partisipasinya dalam e-katalog lokal.
2. Mengevaluasi kualitas klaster yang dihasilkan oleh masing-masing algoritma menggunakan metrik seperti Silhouette Score.
3. Menyajikan hasil segmentasi UMKM sebagai dasar untuk rekomendasi strategi pembinaan yang lebih tepat sasaran.

1.5 Manfaat

Manfaat tugas akhir ini adalah:

1. Memberikan gambaran segmentasi UMKM yang lebih komprehensif melalui perbandingan tiga algoritma clustering, sehingga dapat menunjukkan pola karakteristik usaha dan tingkat partisipasi UMKM dalam e-katalog local.
2. Menjadi bahan pertimbangan bagi pemerintah daerah dalam merumuskan strategi pembinaan UMKM berdasarkan klaster yang terbentuk.
3. Menjadi acuan bagi pelaku UMKM untuk memahami posisi usaha mereka dalam kelompok tertentu serta mendorong peningkatan partisipasi pada e-katalog lokal sebagai peluang memperluas pasar dan akses ke pengadaan pemerintah.
4. Menyediakan referensi akademik terkait penerapan algoritma K-Means Clustering dalam pemetaan UMKM di tingkat daerah.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

BAB 1. PENDAHULUAN

Bab ini berisi uraian umum mengenai latar belakang yang menjelaskan alasan dan urgensi dilakukannya penelitian terkait segmentasi UMKM dalam konteks digitalisasi melalui e-katalog lokal. Selain itu, bab ini memuat identifikasi masalah penelitian, perumusan masalah, batasan penelitian, tujuan dan sasaran penelitian, serta manfaat yang diharapkan baik secara teoritis maupun praktis. Pada bagian akhir bab ini juga dijelaskan sistematika penulisan yang menggambarkan alur penyusunan keseluruhan isi laporan penelitian.

BAB 2. LANDASAN TEORI

Bab ini menyajikan berbagai teori dan konsep yang relevan dengan topik penelitian. Pembahasan mencakup teori dasar mengenai Usaha Mikro, Kecil, dan Menengah (UMKM), digitalisasi pengadaan barang dan jasa melalui e-katalog, serta konsep dan prinsip dari algoritma K-Means Clustering sebagai metode analisis data. Selain itu, bab ini juga menguraikan penelitian-penelitian terdahulu yang menjadi landasan konseptual penelitian ini, termasuk studi-studi yang menerapkan metode clustering pada konteks UMKM atau digitalisasi. Bab ini diakhiri dengan kerangka berpikir (framework) yang menggambarkan hubungan antara teori, variabel penelitian, dan metode analisis yang akan digunakan.

BAB 3. METODOLOGI PENELITIAN

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Bab ini menjelaskan secara rinci pendekatan dan metode penelitian yang digunakan dalam mencapai tujuan penelitian. Pembahasannya meliputi jenis penelitian, sumber dan teknik pengumpulan data, populasi serta sampel data yang dianalisis, dan variabel yang digunakan dalam proses segmentasi. Selain itu, dijelaskan pula tahapan pra-pemrosesan data (data preprocessing), metode analisis menggunakan algoritma K-Means Clustering, serta teknik evaluasi hasil menggunakan Silhouette Score. Pengolahan data dilakukan dengan bantuan perangkat lunak Orange Data Mining untuk memastikan akurasi dan efisiensi dalam proses analisis.

BAB 4. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil pengolahan dan analisis data penelitian mengenai segmentasi UMKM Kota Pekanbaru berdasarkan partisipasi dalam e-Katalog Lokal. Pembahasan mencakup penerapan algoritma K-Means, DBSCAN, dan Hierarchical Clustering serta evaluasi kualitas klaster menggunakan Silhouette Score untuk membandingkan kinerja masing-masing algoritma(Zhou dan Sun, 2024).

BAB 5. PENUUTP

Bab terakhir berisi kesimpulan yang diperoleh berdasarkan hasil penelitian dan pembahasan pada bab sebelumnya. Kesimpulan memuat ringkasan temuan utama yang berkaitan dengan segmentasi UMKM dan tingkat partisipasi dalam e-katalog lokal. Selain itu, bab ini juga menyajikan saran atau rekomendasi yang dapat dijadikan masukan bagi pemerintah daerah, pelaku UMKM, maupun peneliti selanjutnya. Rekomendasi tersebut diharapkan dapat menjadi dasar bagi pengembangan strategi pembinaan UMKM berbasis data serta pengembangan penelitian di bidang data mining dan kebijakan ekonomi digital(Vankayalapati, Ghutugade, Vannapuram, dan Prasanna, 2021).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 2

LANDASAN TEORI

2.1 UMKM

UMKM merupakan kategori usaha yang didefinisikan secara hukum melalui Undang-Undang Nomor 20 Tahun 2008 Pasal 1, dengan parameter pengukuran meliputi total aset, volume omzet tahunan, dan kapasitas tenaga kerja. Dalam realitas praktis, UMKM hadir dalam beragam bentuk dan sektor ekonomi, dari usaha perdagangan tradisional, industri pengolahan skala kecil, hingga jasa-jasa berbasis keahlian lokal. Variasi ini dapat dilihat dalam, yang menampilkan landscape nyata UMKM dengan berbagai karakteristik usaha yang bersifat heterogen namun saling melengkapi dalam struktur perekonomian masyarakat. Gambar 2.1.



Gambar 2.1. Kegiatan bazar UMK

Undang-Undang Republik Indonesia Nomor 20 Tahun 2008 tentang Usaha Mikro, Kecil, dan Menengah (UMKM) memberikan dasar hukum yang jelas mengenai definisi dan kriteria klasifikasi UMKM di Indonesia (UU No.20 Tahun 2008) yang dijelaskan pada tabel berikut Tabel 2.1.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak menggantikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 2.1. Klasifikasi Usaha Berdasarkan Undang-Undang

Kategori	Definisi Menurut Undang-Undang
Usaha Mikro	Usaha produktif milik orang perorangan dan/atau badan usaha perorangan yang memenuhi kriteria usaha mikro sesuai ketentuan Undang-Undang.
Usaha Kecil	Usaha ekonomi produktif yang berdiri sendiri, dilakukan oleh orang perorangan atau badan usaha yang bukan anak perusahaan atau cabang perusahaan dari usaha menengah atau besar, serta memenuhi kriteria usaha kecil menurut Undang-Undang.
Usaha Menengah	Usaha ekonomi produktif yang berdiri sendiri, dilakukan oleh orang perorangan atau badan usaha yang bukan anak perusahaan atau cabang perusahaan dari usaha kecil maupun besar, dengan kekayaan bersih atau hasil penjualan tahunan sesuai ketentuan Undang-Undang.
Usaha Besar	Usaha ekonomi produktif yang dilakukan oleh badan usaha dengan kekayaan bersih atau penjualan tahunan lebih besar dibanding usaha menengah, termasuk BUMN, swasta, patungan, dan usaha asing.
Dunia Usaha	Seluruh usaha yang melakukan kegiatan ekonomi di Indonesia, meliputi usaha mikro, kecil, menengah, dan besar yang berdomisili di Indonesia.

Dalam regulasi tersebut, UMKM dibagi menjadi tiga kategori utama, yaitu usaha mikro, kecil, dan menengah, yang masing-masing dibedakan berdasarkan jumlah aset bersih dan omzet tahunan. Usaha mikro didefinisikan sebagai usaha produktif milik perorangan dengan aset bersih paling banyak Rp50 juta dan omzet tahunan tidak lebih dari Rp300 juta. Usaha kecil adalah usaha produktif yang memiliki aset bersih lebih dari Rp50 juta hingga Rp500 juta dan omzet tahunan lebih dari Rp300 juta hingga Rp2,5 miliar. Sedangkan usaha menengah memiliki aset bersih antara Rp500 juta hingga Rp10 miliar dengan omzet tahunan Rp2,5 miliar sampai Rp50 miliar. Klasifikasi ini penting karena menjadi acuan dalam pemberian fasilitas pembinaan, bantuan permodalan, serta dukungan kebijakan dari pemerintah, baik di tingkat pusat maupun daerah (Suci, 2022). berikut adalah uud UU No.20 Tahun 2008.

Hak Cipta Dilindungi Undang-Undang

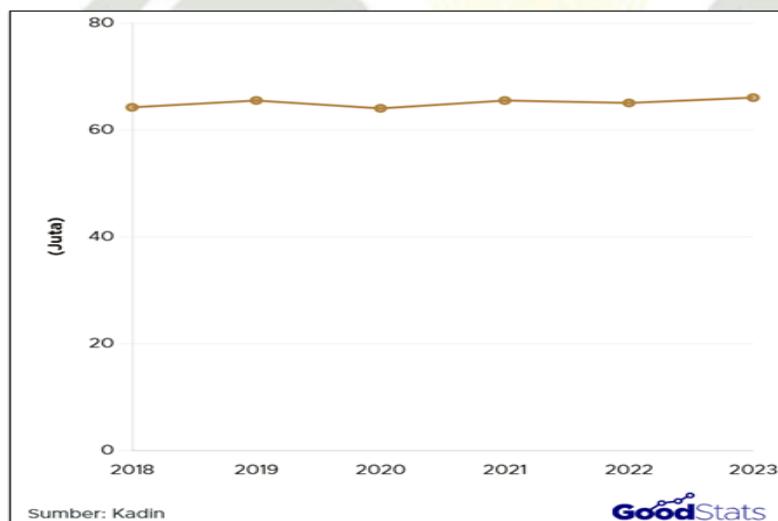
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

- 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

(Husna, 2020) menjelaskan bahwa menurut kebijakan nasional, kriteria usaha mikro, kecil, dan menengah ditetapkan berdasarkan aset bersih serta omzet tahunan. Untuk usaha mikro, kekayaan bersih maksimal ditetapkan sebesar Rp 50.000.000 (tidak termasuk tanah dan bangunan usaha), sedangkan omzet maksimalnya adalah Rp 300.000.000 per tahun. Usaha kecil dibedakan dengan aset lebih dari Rp 50.000.000 sampai dengan Rp 500.000.000 dan omzet antara Rp 300.000.000 hingga Rp 2.500.000.000 per tahun. Sementara usaha menengah memiliki kriteria aset bersih antara Rp 500.000.000 hingga Rp 10.000.000.000, serta omzet tahunan antara Rp 2.500.000.000 sampai maksimal Rp 50.000.000.000 (Husna, 2020)

Perkembangan jumlah UMKM Indonesia dari tahun 2018-2023 Berdasarkan laporan Kadin, pada tahun 2023, jumlah UMKM di Indonesia mencapai 66 juta, yang menunjukkan pertumbuhan sebesar 1,52% dibandingkan tahun sebelumnya. Kontribusi UMKM terhadap Pendapatan Domestik Bruto (PDB) Indonesia mencapai 61%, setara Rp9.580 triliun di tahun tersebut. UMKM juga menyerap sekitar 117 juta pekerja atau sekitar 97% dari total tenaga kerja. yang dapat di lihat pada grafik Gambar 2.2.



Gambar 2.2. Grafik UMKM di Indonesia

Data pertumbuhan pelaku UMKM dalam 5 tahun terakhir menggambarkan perubahan yang fluktuatif, mulai dari 2018 yang mencatatkan total keseluruhan sebesar 64,19 juta, diikuti oleh kenaikan di tahun 2019 yang mencapai 65,47 juta. Namun, pada tahun 2020, terjadi penurunan jumlah UMKM menjadi 64 juta, diikuti kenaikan pada tahun berikutnya menjadi 65,46 juta usaha. Jumlahnya kembali turun

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak menghargai kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

pada 2022 menjadi 65 juta. Terakhir data tahun 2023 mencatatkan kenaikan dengan total 66 juta pelaku UMKM. Kategorisasi UMKM sendiri didasarkan pada besaran modal usaha saat pendirian. Jika modal usaha awal mencapai maksimal Rp1 miliar (tidak termasuk tanah dan bangunan tempat usaha), maka dikategorikan kelas Usaha Mikro. Lalu, usaha dengan modal kisaran Rp1 miliar hingga Rp5 miliar akan masuk dalam kategori Usaha Kecil.

Penjabaran kriteria ini diperkuat oleh penelitian Sinaga (2024) dalam Jurnal Ilmiah AccUsi, yang menekankan bahwa klasifikasi UMKM berbasis aset dan omzet tidak hanya berfungsi sebagai instrumen statistik, tetapi juga sebagai dasar dalam merancang kebijakan pengembangan ekonomi kerakyatan. Menurut Sinaga, pemahaman yang tepat mengenai definisi UMKM akan membantu pemerintah dalam menetapkan program pembinaan, penyaluran bantuan, serta strategi pemberdayaan yang lebih tepat sasaran, terutama di tengah dinamika transformasi digital dan perubahan pasar. Dengan kata lain, pengelompokan yang jelas antara usaha mikro, kecil, dan menengah memberikan kejelasan posisi bagi pelaku usaha dalam ekosistem ekonomi nasional (Sinaga, Martina, dan Purba, 2024).

Sementara itu, pada tingkat internasional, definisi UMKM menunjukkan keragaman yang cukup besar. (*Understanding SME Heterogeneity: Towards Policy Relevant Typologies*, 2021) mencatat bahwa tidak ada standar global yang baku dalam mendefinisikan UMKM, sehingga tiap negara menggunakan indikator yang berbeda sesuai dengan konteks ekonominya. Di negara-negara Eropa, misalnya, SMEs umumnya diklasifikasikan berdasarkan jumlah karyawan, dengan batas ≤ 250 karyawan. Sebaliknya, beberapa negara di Asia dan Afrika lebih menekankan pada omzet tahunan atau total aset sebagai indikator utama. Heterogenitas ini memperlihatkan bahwa meskipun UMKM berperan penting secara global, pendekatan klasifikasinya sangat bergantung pada kebijakan domestik masing-masing negara (*Understanding SME Heterogeneity*, 2021).

Menurut World Bank, usaha kecil dan menengah (SMEs) menyumbang sekitar 90% dari seluruh unit usaha dan lebih dari 50% tenaga kerja di berbagai negara berkembang. McKinsey dalam laporannya *A Microscope on Small Businesses* juga menyebut bahwa di negara maju, UMKM menyumbang sekitar dua pertiga tenaga kerja sektor bisnis, sedangkan di negara berkembang angkanya bisa mendekati empat per lima tenaga kerja sektor bisnis (Adian dkk., 2020). Hal ini menegaskan bahwa UMKM bukan hanya entitas kecil dalam ekonomi, melainkan pilar utama yang menyokong produktivitas, stabilitas, dan inklusi ekonomi global.

Usaha mikro, kecil, dan menengah (UMKM) memegang peranan strategis

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengikuti kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dalam pembangunan ekonomi nasional. Selain berkontribusi pada pertumbuhan ekonomi dan penyerapan tenaga kerja, UMKM juga menjadi pilar ekonomi kerakyatan yang berfungsi mengurangi kemiskinan dan pengangguran, sekaligus mendorong pemerataan hasil pembangunan (Andira dan Chandra, 2023).

UMKM menghadapi tantangan besar dalam proses digitalisasi karena keterbatasan sumber daya, kurangnya tenaga kerja terampil, serta rendahnya literasi teknologi yang menghambat pemanfaatan alat digital secara optimal. Selain itu, keterbatasan akses ke pasar yang lebih luas juga menjadi hambatan utama, di mana banyak UMKM kesulitan berintegrasi dengan platform digital maupun rantai nilai global sehingga daya saing mereka terbatas. Tantangan yang dihadapi tidak hanya bersifat teknologis, tetapi juga struktural, karena digitalisasi menuntut UMKM untuk menata ulang proses bisnis, model usaha, serta strategi dalam menjalin hubungan dengan pelanggan (Faiz, Le, dan Masli, 2024).

2.2 Digitalisasi UMKM dan E-Katalog Lokal

Transformasi digital dan digitalisasi sering digunakan secara bergantian, tetapi secara konseptual berbeda. Digitalisasi merujuk pada penggunaan teknologi digital untuk mengubah proses analog menjadi digital misalnya, mencatat transaksi secara manual menjadi memakai spreadsheet atau sistem kasir elektronik tanpa mengubah model bisnis secara mendasar. Sebaliknya, transformasi digital mencakup perubahan yang lebih mendalam: model usaha baru, proses, interaksi pelanggan, dan struktur organisasi.

Sebagai contoh, dalam studi Digital Transformation of SMEs as an Innovation Process (2024) ditemukan bahwa transformasi digital SMEs melibatkan tiga dimensi utama teknologi, sumber daya manusia, dan organisasi/relasi dan peran strategi eksplisit oleh manajer serta dukungan eksternal sangat krusial (Romero dan Mammadov, 2024). Di konteks Indonesia, model konseptual R-Digital Transformation Model dari Indonesia menunjukkan bahwa transformasi digital di kalangan UMKM selama pandemi dipicu terutama oleh tekanan eksternal seperti pembatasan sosial dan kebutuhan pasar, sementara kesiapan internal (misalnya strategi, kemampuan SDM digital) masih rendah (Muditomo dan Wahyudi, 2020). Dengan memahami perbedaan antara digitalisasi dan transformasi digital, penelitian ini akan mengacu pada transformasi digital sebagai kerangka konseptual untuk menjelaskan bagaimana UMKM tidak hanya mengadopsi teknologi, tetapi mengubah model usaha mereka melalui partisipasi di e-katalog lokal. Transformasi digital merupakan proses melakukan sesuatu dengan cara baru melalui penerapan teknologi digital

untuk merancang ulang bisnis, tidak sekadar otomatisasi atau digitalisasi proses yang ada, melainkan juga mencakup perubahan model usaha, penyesuaian rantai nilai, serta penciptaan produk dan layanan baru. Melalui transformasi ini, UMKM memiliki peluang untuk memperluas akses pasar, terhubung dengan platform digital, dan meningkatkan keterlibatan pelanggan, sehingga daya saing mereka semakin kuat dalam era ekonomi digital, termasuk melalui pemanfaatan E-Katalog Lokal (Kahveci, 2025).

E-Katalog merupakan sistem informasi berbasis elektronik yang digunakan pemerintah untuk memfasilitasi proses pengadaan barang dan jasa secara lebih efisien, transparan, dan akuntabel. Dalam platform ini, produk dan jasa ditampilkan dalam bentuk katalog digital yang berisi informasi lengkap mulai dari spesifikasi teknis, harga, hingga penyedia. Berikut merupakan logo *dari e-Catalogue* Gambar 2.3



Gambar 2.3. Logo E-Katalog

Menurut (Ahmad, 2024), implementasi e-katalog di Kota Pekanbaru berperan penting dalam membuka akses pasar baru bagi UMKM karena proses pengadaan yang sebelumnya tertutup menjadi lebih transparan dan kompetitif, sehingga UMKM dapat lebih mudah ikut serta dalam rantai pasok pemerintah. Hal ini sejalan dengan penelitian (Prikustiawan, 2023) yang mengkaji e-katalog lokal di Surabaya (E-Peken). Mereka menekankan bahwa e-katalog tidak hanya sekadar instrumen digitalisasi, tetapi juga instrumen kebijakan yang mendorong penguatan ekonomi lokal dengan memberi ruang lebih besar bagi UMKM daerah untuk memasarkan produknya dalam pengadaan barang/jasa pemerintah. Dengan demikian, e-katalog

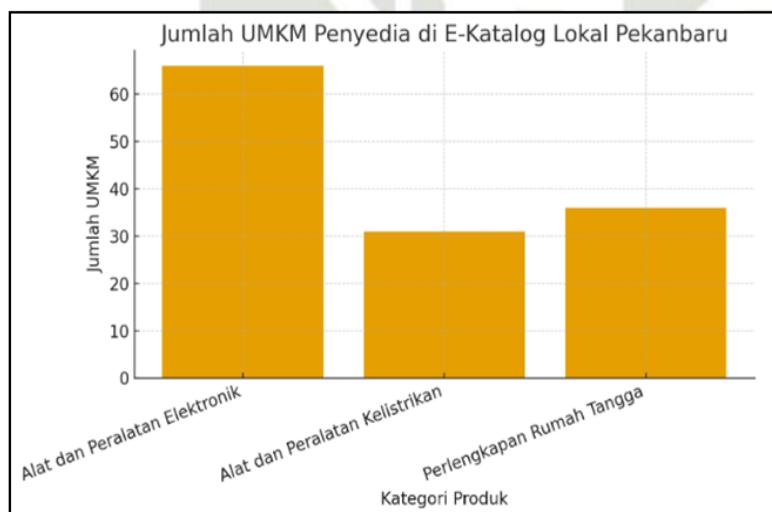
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dapat dipahami sebagai sarana transformasi digital sekaligus kebijakan afirmatif pemerintah untuk memperluas partisipasi UMKM dalam sistem pengadaan.

Katalog elektronik pada dasarnya dibuat untuk mempermudah penjualan produk unggulan perusahaan dalam negeri secara digital. Perkembangannya tidak lagi hanya terbatas pada katalog nasional atau sektoral, tetapi juga mencakup katalog elektronik tingkat lokal. Dalam katalog tersebut, produk ditampilkan dengan informasi lengkap, seperti spesifikasi, fungsi, kinerja, dan harga (Pawe, Utami, dan Man, 2024).

Menurut The Effect of E-Purchasing Implementation Through E-Catalogue on Procurement Performance, e-purchasing yang diatur dalam Peraturan Presiden Republik Indonesia Nomor 12 Tahun 2021 dipahami sebagai mekanisme pembelian barang dan jasa melalui katalog elektronik (e-catalog) maupun toko daring. Pelaksanaannya melalui e-catalogue mencakup beberapa dimensi utama, yaitu efisiensi proses (penghematan biaya dan waktu), transparansi (ketersediaan informasi yang terbuka untuk meminimalisir kecurangan), akurasi dan pengendalian (pemantauan serta pendokumentasian transaksi yang lebih tepat), serta persaingan sehat (akses yang terbuka dan keterlibatan penyedia lokal). Oleh karena itu, e-catalog tidak hanya berperan sebagai sarana digitalisasi pengadaan, tetapi juga sebagai instrumen untuk meningkatkan efisiensi, efektivitas, transparansi, dan akuntabilitas dalam proses pengadaan barang dan jasa pemerintah yang ada dalam Gambar 2.4(Suryani, Nugroho, dan Cadith, 2025).



Gambar 2.4. Garafik jumlah UMKM di Pekanbaru tahun 2023

Partisipasi UMKM dalam program e-katalog masih rendah karena meng-

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengujikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

hadapi sejumlah kendala, seperti minimnya literasi digital yang membuat pelaku usaha belum memahami mekanisme maupun proses pembuatan e-katalog, terbatasnya sosialisasi serta informasi teknis dari pihak terkait, dan keterbatasan akses teknologi maupun kapasitas usaha. Hal ini menyebabkan banyak UMKM kesulitan menyesuaikan diri dengan perkembangan teknologi digital, padahal e-katalog memiliki peran penting dalam memperluas jangkauan pemasaran produk sekaligus meningkatkan peluang UMKM untuk terlibat dalam pengadaan barang dan jasa pemerintah (Setyowati dkk., 2023).

Penelitian yang dilakukan oleh (Pawe dkk., 2024) menunjukkan bahwa optimalisasi e-katalog lokal membutuhkan strategi kebijakan pemerintah daerah yang menyeluruh. Strategi tersebut meliputi sosialisasi kepada penyedia lokal mengenai mekanisme pendaftaran produk, penyediaan program pelatihan dan pendampingan teknis untuk meningkatkan literasi digital UMKM, serta penguatan perencanaan pengadaan agar kebutuhan belanja pemerintah sesuai dengan produk yang tersedia dalam katalog local (Pawe dkk., 2024). Selanjutnya, studi oleh (Prikustiawan, 2023) mengenai implementasi e-katalog “E-Peken” di Kota Surabaya mendeskripsikan bahwa kebijakan e-katalog lokal efektif dalam memperluas akses pasar UMKM dan meningkatkan keterlibatan mereka dalam pengadaan barang/jasa pemerintah (Prikustiawan, 2023). Akan tetapi, hasil penelitian tersebut juga menunjukkan bahwa efektivitas kebijakan masih dipengaruhi oleh keterbatasan kualitas produk, rendahnya pemahaman teknis, dan kebutuhan standardisasi produk agar sesuai dengan persyaratan pengadaan pemerintah. Temuan dari kedua penelitian ini menegaskan bahwa keberhasilan implementasi e-katalog lokal tidak hanya bergantung pada regulasi, tetapi juga pada dukungan teknis serta peningkatan kapasitas UMKM.

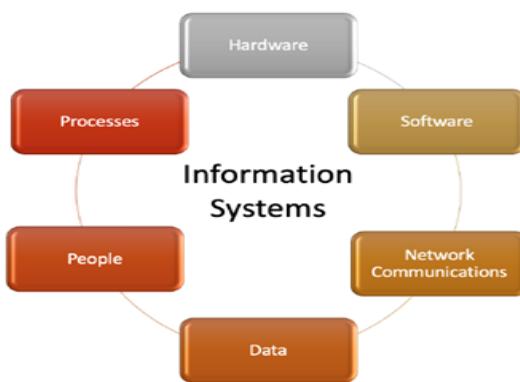
2.3 Data Mining

Menurut LibreTexts (2023), sistem informasi terdiri dari lima komponen utama, yaitu hardware, software, data, people, dan process. Yang ada dalam Gambar 2.5

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak menghargai kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2.5. Komponen sistem informasi

- a) Hardware merujuk pada perangkat fisik seperti komputer, server, dan smartphone yang digunakan untuk menginput, menyimpan, dan mengakses informasi. Dalam konteks UMKM, hardware ini mencakup perangkat kasir elektronik atau laptop yang dipakai untuk mencatat transaksi dan mengelola stok.
- b) Software merupakan perangkat lunak yang menjalankan fungsi sistem, mulai dari aplikasi pengolah data, sistem kasir (point of sales), manajemen inventori, hingga aplikasi e-catalog yang memungkinkan UMKM mengunggah produk ke platform digital pemerintah.
- c) Data menjadi inti dari sistem informasi, meliputi catatan transaksi, stok barang, riwayat pelanggan, hingga harga produk. Data inilah yang diolah untuk menghasilkan informasi yang berguna bagi pengambilan keputusan.
- d) People merujuk pada individu yang berinteraksi dengan sistem, seperti pemilik UMKM, staf administrasi, maupun operator katalog yang bertugas mengelola input dan memastikan kelancaran alur informasi. Terakhir, process adalah prosedur atau mekanisme yang mengatur bagaimana data dikumpulkan, diproses, divalidasi, dan disajikan kembali dalam bentuk laporan yang siap digunakan untuk analisis bisnis.
- e) communication networks adalah jaringan komunikasi yang memungkinkan pertukaran data antar perangkat dan sistem. Dalam konteks UMKM, komponen ini sangat penting karena koneksi internet menjadi prasyarat untuk mengakses e-catalog lokal, memperbarui produk secara daring, serta menjangkau konsumen lebih luas. Dengan adanya jaringan, sistem informasi UMKM tidak hanya berfungsi secara internal, tetapi juga terhubung dengan ekosistem digital yang lebih luas.
- f) Dalam praktiknya, kelima komponen ini saling terintegrasi. Misalnya, seo-

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak menghargai kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

rang pemilik UMKM menggunakan laptop (hardware) dengan aplikasi manajemen penjualan (software) untuk mencatat transaksi (data), yang kemudian diverifikasi oleh staf (people), sebelum akhirnya diolah menjadi laporan penjualan harian sebagai dasar keputusan bisnis (process). Dengan cara ini, sistem informasi tidak hanya berfungsi sebagai alat pencatatan, tetapi juga sebagai sarana strategis yang meningkatkan efisiensi, transparansi, dan daya saing UMKM di era digital (Pham, Desai-Naik, Hammond, dan Abdeljabbar, 2021)

Sistem Informasi Manajemen berperan dalam mengolah data menjadi informasi yang akurat dan tepat waktu guna mendukung pengambilan keputusan, sekaligus mempermudah fungsi perencanaan, pengendalian, serta operasional organisasi agar lebih efisien. Informasi yang dihasilkan dari pengolahan data tersebut menjadi sumber daya yang sangat vital, setara dengan modal, tenaga kerja, dan mesin, karena tanpa informasi, keberlangsungan organisasi modern akan sulit terjaga (Sidhu dan Sharma, 2023).

Sistem informasi berfungsi sebagai fondasi utama dalam pengelolaan data pada organisasi maupun UMKM. Data yang dihasilkan melalui sistem informasi, baik berupa transaksi, inventori, maupun catatan manajerial, menjadi bahan baku yang sangat penting untuk mendukung pengambilan keputusan. Agar data tersebut memiliki nilai strategis, diperlukan pendekatan analitis yang lebih mendalam. Di sinilah konsep Business Intelligence (BI) berperan, yakni seperangkat proses, teknologi, dan arsitektur yang mengubah data mentah menjadi informasi serta pengetahuan yang dapat digunakan untuk pengambilan keputusan.

Salah satu komponen penting dalam BI adalah data mining, yang didefinisikan sebagai proses komputasional untuk mengekstraksi pengetahuan, menemukan pola, hubungan, maupun tren yang tersembunyi dalam data yang besar. Dalam kerangka kerja Knowledge Discovery in Databases (KDD), data mining menempati posisi inti karena berfokus pada transformasi data menjadi pengetahuan yang bermanfaat. Dengan demikian, keterkaitan antar konsep ini dapat dipahami melalui alur: sistem informasi menghasilkan data → data dianalisis melalui BI → data mining mengekstrak pola → hasilnya menjadi pengetahuan baru untuk mendukung pengambilan keputusan.

Model konseptual ini sejalan dengan pandangan (Risco-Ramos dkk., 2023), yang menjelaskan bahwa data mining merupakan subdomain dari artificial intelligence yang bertujuan menghasilkan pengetahuan dari data, dengan cara menemukan pola, relasi, maupun tren yang relevan bagi pengguna. Pengetahuan yang dihasilkan tidak hanya bersifat deskriptif, tetapi juga dapat digunakan sebagai

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak menghargai kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dasar prediksi dan perencanaan strategis. Oleh karena itu, dalam konteks penelitian ini, sistem informasi berperan sebagai penyedia data, sedangkan data mining berfungsi mengubah data tersebut menjadi informasi bernilai guna yang dapat mendukung kebijakan dan strategi pengembangan UMKM melalui pemanfaatan e-katalog lokal(Risco-Ramos dkk., 2023)

Data mining merupakan serangkaian teknik komputasi yang digunakan untuk menemukan pola tersembunyi, korelasi, maupun anomali dalam kumpulan data yang besar. Metode seperti klasifikasi, klasterisasi, association rule mining, dan deteksi anomali menjadi elemen penting dalam meningkatkan kemampuan Decision Support System (DSS). Selain itu, penerapan algoritma machine learning seperti decision tree, neural network, dan support vector machine memungkinkan DSS meningkatkan akurasi prediksi serta beradaptasi dengan perubahan pola data. Berkat perkembangan ini, DSS mampu memberikan prediksi yang lebih tepat, wawasan yang dapat langsung ditindaklanjuti, serta dukungan pengambilan keputusan otomatis di berbagai sektor, termasuk keuangan, kesehatan, pemasaran, hingga administrasi public (Munawaroh, 2023).

2.4 Clustering

Klasterisasi atau clustering adalah metode untuk mengelompokkan data. Proses ini dilakukan dengan membagi sekumpulan objek data ke dalam beberapa himpunan yang disebut klaster. Tujuan utamanya adalah menempatkan data atau objek ke dalam suatu klaster sehingga setiap kelompok berisi data yang memiliki tingkat kemiripan setinggi mungkin (Fikri, 2024)

Clustering atau klasterisasi merupakan salah satu teknik utama dalam data mining yang bertujuan mengelompokkan sekumpulan objek data ke dalam beberapa himpunan (cluster) berdasarkan tingkat kesamaan di antara objek tersebut. Inti dari clustering adalah prinsip homogenitas internal (data dalam satu cluster harus memiliki tingkat kesamaan yang tinggi) dan heterogenitas eksternal (data antar cluster harus berbeda secara signifikan). Dengan kata lain, clustering berusaha menemukan struktur alami dalam data tanpa perlu label atau kategori yang sudah ditentukan sebelumnya (Syaputri et al., 2021).

Sebagai metode eksplorasi data, clustering memiliki peran penting dalam fluktuasi kumpulan data yang besar dan kompleks menjadi kelompok-kelompok yang lebih mudah dipahami. Proses ini memungkinkan peneliti atau pengambil keputusan untuk menemukan pola tersembunyi, tren, maupun anomali yang tidak terlihat secara kasat mata. Misalnya, dalam bidang pemasaran, clustering

digunakan untuk segmentasi konsumen berdasarkan perilaku belanja (Afrizal dkk., 2025). Dalam bidang kesehatan, clustering dimanfaatkan untuk mengelompokkan pasien berdasarkan gejala atau riwayat medis, yang kemudian membantu diagnosis dan rencana perawatan. Sementara itu, dalam bidang pendidikan, clustering sering digunakan untuk mengelompokkan siswa berdasarkan kemampuan belajar bahkan dalam keamanan siber berperan dalam mendeteksi aktivitas abnormal pada sistem jaringan (H. Ramadhan, Abdan Kamaludin, Nasrullah, dan Rolliawati, 2023)

Dalam ranah pembelajaran mesin, clustering digolongkan sebagai metode pembelajaran tanpa pengawasan, karena tidak memerlukan data latih dengan label. Hal ini berbeda dengan klasifikasi yang merupakan pembelajaran terbimbing, dimana model membutuhkan data berlabel untuk memprediksi kategori data baru. Karakteristik unsupervised inilah yang membuat clustering sangat relevan untuk penelitian yang fokus pada eksplorasi pola data, seperti segmentasi UMKM yang tidak memiliki label kategori sebelumnya(Madhulatha, 2021).

Secara umum, metode clustering dapat dibedakan dalam beberapa pendekatan utama (Sajana, Sheela Rani, dan Narayana, 2023):

1. Metode partisi → membagi dataset ke dalam sejumlah cluster yang telah ditentukan sebelumnya, contohnya K-Means dan K-Medoids. Metode ini cocok untuk data berukuran besar dan menghasilkan cluster yang relatif jelas.
2. Metode hierarki → menghasilkan struktur hierarki berbentuk dendrogram yang menunjukkan hubungan antar cluster. Metode ini cocok untuk analisis eksploratif yang memerlukan gambaran hubungan antar kelompok data.
3. Metode berbasis kepadatan → seperti DBSCAN, yang membentuk cluster berdasarkan kepadatan data. Metode ini unggul dalam menemukan cluster dengan bentuk yang tidak beraturan dan mampu mendeteksi outlier (Radicic dan Petković, 2023)
4. Metode berbasis model → misalnya Gaussian Mixture Model (GMM), yang menggunakan pendekatan probabilistik dengan asumsi bahwa data berasal dari kombinasi distribusi statistik tertentu.

Setiap metode memiliki kelebihan dan kelemahan. Partisi seperti K-Means sederhana dan efisien, tetapi memerlukan jumlah cluster yang ditentukan sejak awal. Pengelompokan hierarki mudah dipahami, tetapi kurang efisien untuk kumpulan data besar. Berbasis kepadatan seperti DBSCAN lebih fleksibel dalam bentuk cluster, tetapi kurang optimal untuk data dengan kepadatan yang bervariasi. Oleh karena itu, pemilihan metode harus disesuaikan dengan karakteristik data dan tujuan

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, pengelompokan data, dan penyajian suatu masalah.
 - b. Pengutipan tidak menghargai kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

penelitian (Madhulatha, 2021)(Sajana dkk., 2023) Selain pemilihan algoritma, kualitas hasil clustering juga perlu dievaluasi. Dengan metode Silhouette Score untuk mengukur seberapa baik data ditempatkan dalam cluster yang tepat, serta Davies-Bouldin Index (DBI) untuk menilai tingkat kemiripan antar cluster (Afrizal dkk., 2025) Evaluasi ini penting agar hasil klasterisasi tidak hanya terbentuk secara teknis, tetapi juga memiliki validitas ilmiah. Dalam konteks penelitian ini, clustering sangat relevan karena UMKM di Kota Pekanbaru memiliki karakteristik beragam dalam hal partisipasi pada e-katalog lokal. Dengan menerapkan teknik clustering, khususnya K-Means, data UMKM dapat dikumpulkan ke dalam kategori seperti UMKM aktif, UMKM baru bergabung, dan UMKM yang belum berpartisipasi. Hasil segmentasi ini akan

Memberikan wawasan yang lebih terarah bagi pemerintah daerah dalam merancang kebijakan pelatihan dan strategi digitalisasi UMKM, sehingga manfaat e-katalog lokal dapat lebih dirasakan oleh seluruh pelaku usaha (Arrahmi et al., 2025).

2.5 Algoritma K-Means Clustering

Sejarah perkembangan algoritma K-Means tidak lepas dari kontribusi beberapa peneliti sejak dekade 1950-an. Hugo Steinhaus (1956) menjadi salah satu yang pertama kali memperkenalkan ide tentang Pembagian objek atau “corps matériels” ke dalam beberapa bagian melalui metode matematis. Gagasan ini digunakan dalam antropologi dan industri sebagai cara membagi data yang heterogen menjadi kelompok yang lebih homogen. Beberapa tahun kemudian, Stuart Lloyd (1957) dari Bell Laboratories mengembangkan pendekatan serupa dalam artikelnya Least Squares Quantization di PCM, yang awalnya dirancang untuk komunikasi sinyal dan elektronika. Algoritma Lloyd fokus pada kuantisasi sinyal acak multidimensi untuk sistem pulse-code modulation, dan meskipun pada saat itu tidak disebut sebagai K-Means, prinsip iterasi minimisasi jarak kuadrat sudah menjadi landasannya (Pérez-Ortega dkk., 2019).

Istilah K-Means secara resmi diperkenalkan oleh James MacQueen pada tahun 1967 melalui artikelnya Beberapa Metode Klasifikasi dan Analisis Pengamatan Multivariat. MacQueen mendefinisikan metode ini sebagai algoritma partisi data ke dalam sejumlah klaster yang meminimalkan varians intra-klaster. Dalam perkembangannya, metode ini juga dikenal dengan berbagai istilah lain, seperti metode pengelompokan dinamis, pengelompokan jarak minimum berulang, pengurutan centroid terdekat, dan h-means. Dengan demikian, meskipun K-Means baru

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak menghargai kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

populer sejak MacQueen memperkenalkannya, akar sejarahnya dapat ditelusuri hingga Steinhaus (1956) dan Lloyd (1957) yang masing-masing meletakkan dasar konsep partisi data multidimensi dan iterasi kuantisasi sinyal (Pérez-Ortega dkk., 2019).

Algoritma k-means bekerja dengan menghitung fungsi jarak. Prosesnya terdiri dari dua tahap, yaitu tahap assignment yang menempatkan setiap titik ke dalam klaster berdasarkan jarak terpendek ke pusat klaster (centroid), serta tahap update yang menghitung ulang posisi centroid setiap klaster berdasarkan hasil pengelompokan titik. Kedua tahap ini dilakukan berulang-ulang hingga posisi centroid tidak lagi berubah (Xia dkk., 2021).

Algoritma K-Means bekerja dengan menentukan jumlah klaster (k), lalu memilih k centroid awal secara acak. Setiap data kemudian dihitung jaraknya, umumnya menggunakan jarak Euclidean, dan ditempatkan ke centroid terdekat. Setelah itu, posisi centroid diperbarui dengan menghitung rata-rata dari data yang ada dalam setiap klaster. Proses penugasan data ke klaster dan pembaruan centroid dilakukan secara berulang (iterasi) hingga posisi centroid tidak banyak berubah lagi, sehingga terbentuk kelompok data yang homogen di dalam klaster dan heterogen antar klaster (Vankayalapati dkk., 2021). Berikut adalah Langkah langkah dalam metode k-means:

2.5.1 Menentukan jumlah cluster yang mau di bentuk

Langkah pertama dalam proses klasterisasi adalah menentukan berapa banyak cluster (kelompok) yang akan dibentuk dari data. Jumlah cluster ini dilambangkan dengan huruf K. Pemilihan nilai K sangat penting karena akan memengaruhi hasil pengelompokan data terlalu sedikit cluster dapat menyebabkan informasi penting terabaikan, sedangkan terlalu banyak cluster dapat membuat hasil sulit diinterpretasikan.

Untuk menentukan jumlah cluster yang optimal, digunakan Metode Silhouette Coefficien digunakan untuk mengukur seberapa baik setiap objek cocok dengan cluster-nya sendiri dibandingkan dengan cluster lainnya. Nilai silhouette score berkisar antara -1 hingga 1. Semakin mendekati 1 berarti pengelompokan semakin baik.

2.5.2 Menentukan centeroid awal secara acak

Setelah jumlah cluster (K) ditentukan, langkah berikutnya adalah menentukan centroid awal, yaitu titik pusat dari masing-masing cluster. Centroid ini berfungsi sebagai acuan awal untuk menghitung jarak antara data dan setiap cluster

- Hak Cipta Diilngungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
 - b. Pengutipan tidak mengutip sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dalam proses iterasi algoritma.

Pada tahap ini, K titik centroid awal dipilih secara acak dari kumpulan data. Setiap centroid mewakili posisi awal dari satu cluster. Pemilihan secara acak ini dilakukan untuk memastikan bahwa setiap cluster memiliki peluang yang sama dalam proses pembentukan kelompok awal.

Dalam beberapa kasus, untuk meningkatkan stabilitas hasil, digunakan metode inisialisasi yang lebih baik seperti K-Means++ Initialization yang memilih titik centroid awal dengan mempertimbangkan jarak antar data, sehingga hasil klasifikasi lebih stabil dan konvergen lebih cepat.

Setelah centroid awal ditetapkan, proses klasterisasi akan dilanjutkan dengan menghitung jarak setiap data terhadap centroid (biasanya menggunakan jarak Euclidean), lalu mengelompokkan data ke dalam cluster terdekat.

2.5.3 Menghitung jarak setiap data ke masing-masing centroid dari masing-masing Clustering

K-Means menggunakan jarak Euclidean sebagai metrik dasar untuk menentukan seberapa dekat suatu titik data ke masing-masing centroid. Dalam metode ini, Jarak antara data x dan centroid μ dihitung menggunakan rumus(Gu dkk., 2009):

$$d(x, \mu) = \sqrt{\sum_{j=1}^m (x_j - \mu_j)^2} \quad (2.1)$$

Keterangan:

- $d(x, \mu)$: jarak Euclidean antara titik data x dan centroid μ
- x_j : nilai atribut ke- j dari titik data x
- μ_j : nilai atribut ke- j dari centroid μ
- m : jumlah dimensi atau variabel pada data

2.5.4 Fungsi Tujuan

K-Means dapat dirumuskan sebagai *Sum of Squared Error* (SSE) atau *Within-Cluster Sum of Squares* (WCSS) yang bertujuan untuk meminimalkan total jarak kuadrat antara data dan pusat klasternya (Pourahmad dkk., 2020), yang dirumuskan sebagai berikut:

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (2.2)$$

Keterangan:

- J : nilai fungsi tujuan (*loss function*) yang ingin diminimalkan
- K : jumlah klaster yang ditentukan
- C_k : himpunan data yang termasuk dalam klaster ke- k
- x_i : titik data ke- i
- μ_k : centroid dari klaster ke- k
- $\|x_i - \mu_k\|^2$: jarak kuadrat Euclidean antara titik data x_i dan centroid μ_k

2.5.5 Memperbarui Centroid

Centroid μ_k dari klaster C_k dihitung ulang sebagai rata-rata seluruh titik data yang berada dalam klaster tersebut, yang dirumuskan sebagai berikut:

$$\mu_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i \quad (2.3)$$

Keterangan:

- μ_k : centroid baru dari klaster ke- k
- $|C_k|$: jumlah titik data dalam klaster ke- k
- $\sum_{x_i \in C_k} x_i$: penjumlahan seluruh vektor data x_i yang berada dalam klaster ke- k

Dalam studi On the Efficiency of K-Means Clustering, penulis menyebutkan bahwa centroid diperbarui dengan membagi vektor jumlah (sum vector) cluster dengan jumlah titik dalam cluster (S. Wang, Sun, dan Bao, 2020).

Proses penugasan (menugaskan titik ke centroid terdekat) dan update centroid ini dilakukan secara iteratif sampai perubahan menjadi sangat kecil atau tidak ada lagi perubahan, sehingga fungsi tujuan mencapai titik minimal lokal (Comprehensive Review of K-Means)(Oti, Olusola, Eze, dan Enogwe, 2021).

Algoritma K-Means banyak digunakan karena sederhana, efisien, dan mudah diimplementasikan, serta bekerja baik pada data dengan klaster yang kompak dan terpisah jelas, bahkan mampu menangani dataset besar. Namun, algoritma ini juga memiliki kelemahan, antara lain harus menentukan jumlah klaster (k) sejak awal, sensitif terhadap posisi awal centroid, mudah dipengaruhi oleh outlier maupun noise, serta kurang efektif pada data dengan bentuk klaster non-sferis atau kepadatan yang bervariasi (Ahmed, Seraj, dan Islam, 2020)

2.6 Aggoritma DBSCAN

2.6.1 Konsep Density-Based Clustering

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) merupakan salah satu algoritma clustering yang berorientasi pada kepadatan data (density-based). Algoritma ini diperkenalkan oleh Ester, Kriegel, Sander, dan Xu (1996) sebagai solusi terhadap keterbatasan metode clustering konvensional seperti K-Means yang hanya mampu membentuk cluster berbentuk bola (spherical) dan sensitif terhadap outlier.

Pada pendekatan density-based, suatu cluster didefinisikan sebagai wilayah dalam ruang data yang memiliki kepadatan tinggi dan dipisahkan oleh area dengan kepadatan lebih rendah. Dengan demikian, DBSCAN berupaya mencari area berdensitas tinggi yang saling terhubung, kemudian membentuknya sebagai sebuah cluster yang utuh. Pendefinisian ini mengacu pada konsep density reachability dan density connectivity sebagaimana dikemukakan oleh (Ester, Kriegel, dan Xu, 1996).

Pendekatan berbasis kepadatan ini menjadikan DBSCAN sangat efektif dalam menemukan pola kompleks pada data nyata yang umumnya memiliki distribusi tidak beraturan. Penelitian kontemporer seperti (Simbolon dan Friskila, 2024) menunjukkan bahwa metode ini mampu mengidentifikasi struktur cluster yang tidak dapat ditangkap oleh pendekatan partisional tradisional, terutama pada dataset dengan tingkat variasi tinggi.

2.6.2 Parameter Eps Dan Minpts

Kinerja dan hasil pengelompokan menggunakan algoritma DBSCAN sangat dipengaruhi oleh dua parameter utama, yaitu ϵ (*epsilon* atau *eps*) dan *MinPts*. Kedua parameter ini berperan penting dalam menentukan struktur klaster yang terbentuk.

Parameter ϵ merupakan radius maksimum yang digunakan untuk menentukan apakah suatu titik termasuk ke dalam lingkungan (*neighborhood*) titik lainnya. Nilai ϵ mendefinisikan jarak maksimum antar dua titik agar dapat dianggap saling berhubungan secara berbasis kepadatan.

Parameter *MinPts* menunjukkan jumlah minimum titik data yang harus berada dalam radius ϵ untuk membentuk suatu wilayah dengan kepadatan tinggi (*dense region*).

- Jika jumlah titik tetangga $\geq MinPts$, maka titik tersebut dikategorikan sebagai *core point*.
- Jika jumlah titik tetangga $< MinPts$, maka titik tersebut tidak termasuk sebagai *core point*.

Ester et al. (1996) menegaskan bahwa pemilihan parameter ϵ dan $MinPts$ sangat memengaruhi kualitas identifikasi klaster. Nilai ϵ yang terlalu kecil cenderung menghasilkan banyak data yang diklasifikasikan sebagai *noise*, sedangkan nilai ϵ yang terlalu besar dapat menyebabkan klaster yang seharusnya terpisah bergabung menjadi satu. Penelitian oleh (H. Ramadhan dkk., 2023) juga menekankan pentingnya penentuan parameter yang optimal untuk menghindari terjadinya *under-clustering* maupun *over-clustering*.

2.6.3 Tipe-Tipe Titik: Core Border, Noise

Proses kerja DBSCAN berlangsung melalui tahapan-tahapan sebagai berikut:

1. Memilih sebuah titik yang belum dikunjungi.
2. Mengukur jumlah tetangga pada radius ϵ .
3. Jika jumlahnya $\geq MinPts$, titik tersebut menjadi core point dan memulai pembentukan cluster.
4. Algoritma kemudian melakukan perluasan cluster melalui titik-titik tetangga yang juga memenuhi syarat densitas.
5. Proses ekspansi berlanjut hingga tidak ada lagi titik dalam wilayah densitas tinggi yang dapat ditambahkan ke cluster.

Mekanisme ini disebut proses density expansion, yang memungkinkan DBSCAN membentuk cluster dengan bentuk yang sangat beragam (non-linear, bercabang, maupun melengkung). Hal ini sejalan dengan temuan (Ester dkk., 1996) serta diperkuat oleh (Simbolon dan Friskila, 2024) yang menegaskan bahwa kemampuan ekspansi berdasarkan keterhubungan densitas merupakan keunggulan teoritis utama DBSCAN.

2.6.4 Mekanisme Pembentukan Cluster

Proses kerja algoritma DBSCAN dimulai dengan memilih satu titik data yang belum pernah dikunjungi. Titik ini kemudian diberi penanda sebagai “visited” untuk memastikan bahwa ia tidak akan dievaluasi kembali pada langkah berikutnya. Setelah titik awal ditentukan, algoritma menghitung seluruh titik lain yang berada dalam radius tertentu (ϵ atau epsilon) dari titik tersebut. Radius ϵ berfungsi sebagai batasan ruang untuk mengidentifikasi tetangga terdekat, sementara jumlah tetangga dihitung berdasarkan jarak Euclidean atau metrik jarak lain yang sesuai dengan karakteristik data. Himpunan titik-titik dalam radius ϵ ini disebut sebagai neighborhood.

Ketika jumlah tetangga dalam Radius ϵ mencapai nilai minimum yang telah

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak menghargai kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

ditentukan (MinPts), titik tersebut dikategorikan sebagai core point. Sebaliknya, apabila jumlah tetangga lebih sedikit dari MinPts, maka titik tersebut sementara dianggap sebagai noise, meskipun status ini dapat berubah apabila kemudian ditemukan sebagai bagian dari wilayah densitas titik inti lainnya. Jika titik memenuhi syarat sebagai core point, DBSCAN akan memulai pembentukan klaster dengan menetapkan titik inti tersebut sebagai pusat klaster awal.

Proses pembentukan klaster dilanjutkan dengan memasukkan semua titik dalam neighborhood ke dalam klaster yang sedang dibentuk, kemudian menambahkan titik-titik tersebut ke dalam daftar antrian ekspansi atau seed list. Selanjutnya, algoritma mengambil satu per satu titik dari daftar tersebut untuk diperiksa. Apabila salah satu titik dalam daftar ekspansi juga memenuhi syarat sebagai core point (memiliki jumlah tetangga yang cukup dalam Radius ϵ), maka neighborhood dari titik tersebut akan digabungkan dengan klaster, dan titik-titik baru yang belum pernah dikunjungi akan dimasukkan ke dalam antrian ekspansi. Mekanisme ini memungkinkan DBSCAN untuk memperluas klaster secara iteratif hingga seluruh titik dalam wilayah densitas tinggi saling terhubung melalui konsep density reachability.

Ekspansi klaster akan berhenti secara otomatis ketika tidak ada lagi titik dalam antrian ekspansi yang dapat diperluas. Setelah satu klaster selesai dibentuk, DBSCAN kembali memilih titik lain yang belum dikunjungi dan mengulangi proses yang sama. Titik yang tidak memenuhi syarat densitas akan tetap berstatus noise, sedangkan titik yang memenuhi syarat akan membentuk klaster baru. Dengan demikian, keseluruhan proses DBSCAN berlangsung secara berulang hingga seluruh titik dalam dataset diproses, dan hasil akhirnya berupa himpunan beberapa klaster yang terbentuk berdasarkan kepadatan serta sejumlah titik yang dikategorikan sebagai noise atau outlier.

Mekanisme ini disebut proses density expansion, yang memungkinkan DBSCAN membentuk cluster dengan bentuk yang sangat beragam (non-linear, bercabang, maupun melengkung). Hal ini sejalan dengan temuan (Ester dkk., 1996) serta diperkuat oleh (Simbolon dan Friskila, 2024) yang menegaskan bahwa kemampuan ekspansi berdasarkan

2.7 Hierarchical Clustering

2.7.1 Hierarchical Clustering

Hierarchical clustering adalah metode pengelompokan yang membentuk struktur bertingkat (hierarchy) antar objek sehingga hasil akhir dapat direpresentasikan sebagai pohon berakar (dendrogram). Metode ini berguna untuk men-

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengurangi kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

gungkap struktur multilevel dalam data misalnya sub segmen di dalam segmen yang seringkali berguna dalam analisis sosial-ekonomi seperti segmentasi UMKM. Dua pendekatan utama adalah agglomerative (bottom-up) dan divisive (top-down). Keunggulan utama hierarchical clustering adalah kemampuannya menunjukkan hubungan antar cluster pada berbagai level pemotongan (cutting level), sehingga peneliti memiliki fleksibilitas dalam memilih jumlah cluster berdasarkan konteks aplikatif.

2.7.1 Konsep Agglomerative dan Divisive

Agglomerative (bottom-up) Metode agglomerative memulai proses dengan menganggap setiap objek sebagai cluster tersendiri (n cluster untuk n objek). Secara iteratif, pasangan cluster yang paling mirip (atau terdekat) digabungkan menjadi cluster yang lebih besar sampai tersisa satu cluster yang memuat seluruh objek. Prosedur ini menghasilkan pohon hierarki dari bawah ke atas.

Divisive (top-down) Sebaliknya, pendekatan divisive memulai dengan satu cluster yang memuat seluruh objek, lalu secara bertahap memecah cluster menjadi subcluster sampai setiap objek berdiri sendiri atau kondisi penghentian terpenuhi. Pendekatan divisive kurang lazim diperlakukan karena biaya komputasinya biasanya lebih tinggi.

Konsep kesamaan/ketidaksamaan Kedua pendekatan bergantung pada ukuran kesamaan (similarity) atau ketidaksamaan (dissimilarity) antar objek/cluster. Ukuran ini biasanya berupa metrik jarak (mis. Euclidean, Manhattan) atau metrik campuran (mis. Gower untuk data numerik + kategorikal). Pemilihan metrik ini menjadi dasar keputusan penggabungan/PEMECATAN cluster pada setiap langkah.

2.7.2 Linkage Methods

Metode *linkage* menentukan cara perhitungan jarak antar klaster pada setiap tahap penggabungan dalam algoritma *Hierarchical Clustering*. Berikut ini disajikan formulasi matematis serta karakteristik utama dari masing-masing metode *linkage*.

1. Single Linkage

$$d_{SL}(A, B) = \min_{x \in A, y \in B} d(x, y) \quad (2.4)$$

Menggabungkan dua cluster berdasarkan sepasang titik terdekat. Cenderung menghasilkan chain effect (cluster memanjang).

2. Complete Linkage

$$d_{CL}(A, B) = \max_{x \in A, y \in B} d(x, y) \quad (2.5)$$

Complete linkage mempertimbangkan jarak terjauh antar pasangan titik

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak menghargai kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dari dua klaster. Pendekatan ini menghasilkan klaster yang lebih kompak(*compact*) dan cenderung meminimalkan diameter klaster.

3. Average Linkage

$$d_{AL}(A, B) = \frac{1}{|A||B|} \sum_{x \in A} \sum_{y \in B} d(x, y) \quad (2.6)$$

Average linkage menggunakan rata-rata jarak seluruh pasangan titik dari dua klaster. Metode ini lebih stabil dibandingkan single linkage dan tidak terlalu sensitif terhadap outlier ekstrem.

4. Ward's Method

Ward menggabungkan dua klaster yang menghasilkan peningkatan terkecil pada jumlah kuadrat deviasi internal (*total within-cluster sum of squares*). Jika $SSE(C)$ menyatakan jumlah kuadrat deviasi dalam klaster C , maka metode Ward memilih pasangan klaster A dan B yang meminimalkan nilai:

$$\Delta(A, B) = SSE(A \cup B) - SSE(A) - SSE(B) \quad (2.7)$$

Secara praktis, Ward cenderung menghasilkan cluster berukuran seimbang dan sering dianjurkan untuk data numerik ketika tujuan adalah meminimalkan variansi internal.

- Single linkage baik untuk menemukan struktur koneksi tetapi rentan terhadap noise dan rantai.
- Complete linkage dan Ward lebih baik untuk menghasilkan cluster kompak.
- Average linkage memberikan kompromi antara single dan complete.
- Pemilihan linkage harus mempertimbangkan tujuan analisis (mis. interpretabilitas vs. sensitivitas terhadap outlier).

2.7.3 Dendrogram

Dendrogram adalah representasi grafis hirarki yang menggambarkan urutan penggabungan atau pemecahan cluster. Pada umumnya:

- Sumbu horizontal merepresentasikan objek (atau label objek).
- Sumbu vertikal menunjukkan jarak atau nilai ketidaksamaan pada saat penggabungan. Tingkat pemotongan pada sumbu vertikal menentukan jumlah cluster yang dihasilkan.

Interpretasi dan pemilihan jumlah cluster:

- *Cutting the dendrogram* pada ketinggian tertentu menghasilkan pemisahan cluster; pemotongan biasanya dilakukan pada lokasi loncatan besar di ketinggian penggabungan (gap yang signifikan menunjukkan pemisahan alami).
- *Cophenetic correlation coefficient* dapat dipakai untuk mengukur seberapa baik dendrogram mempertahankan informasi jarak asli:

$$c = \frac{\sum_{i < j} (d_{ij} - \bar{d})(t_{ij} - \bar{t})}{\sqrt{\sum_{i < j} (d_{ij} - \bar{d})^2 \sum_{i < j} (t_{ij} - \bar{t})^2}} \quad (2.8)$$

di mana d_{ij} merupakan jarak asli antara objek ke- i dan ke- j , sedangkan t_{ij} adalah jarak cophenetic, yaitu jarak pada dendrogram saat objek i dan j pertama kali digabungkan. Simbol garis atas (\bar{d} dan \bar{t}) menyatakan nilai rata-rata dari masing-masing jarak tersebut. Nilai koefisien cophenetic c yang mendekati 1 menunjukkan bahwa struktur dendrogram merepresentasikan jarak asli data dengan baik. Visualisasi tambahan:

- Heatmap matriks jarak dikombinasikan dengan dendrogram dapat membantu menilai struktur cluster.
- Dendrogram memungkinkan identifikasi sub-cluster yang mungkin relevan untuk kebijakan lokal (mis. kelompok UMKM yang membutuhkan pelatihan khusus).

2.7.4 Kelebihan Dan Kelemahan Kelebihan

Kelebihan

1. Menampilkan struktur hirarkis memberikan wawasan multilevel (mis. sub-segmentasi UMKM menurut omzet \rightarrow produk \rightarrow partisipasi e-katalog).
2. Tidak memerlukan nilai K awal jumlah cluster dapat ditentukan melalui interpretasi dendrogram.
3. Fleksibilitas linkage dapat disesuaikan dengan tujuan analisis (memprioritaskan compactness, connectivity, atau keseimbangan ukuran cluster).
4. Interpretabilitas dendrogram intuitif bagi pembuat kebijakan untuk menelusuri tingkat kemiripan antar kelompok UMKM.
5. Cocok untuk dataset kecil–menengah memberikan hasil yang stabil dan mudah dianalisis.

Kelemahan

1. Sensitif terhadap outlier terutama pada metode single/complete linkage; outlier dapat mengubah struktur penggabungan.
2. Irreversibilitas penggabungan setelah dua cluster digabung, keputusan tidak dapat dibatalkan sehingga kesalahan awal dapat mempengaruhi hasil akhir.
3. Pemilihan metrik jarak hasil sangat tergantung pada metrik yang digunakan; untuk data campuran (numerik + kategorikal) perlu metrik khusus (mis. Gower) atau encoding yang cermat.
4. Skala variabel variabel dengan skala besar akan mendominasi jarak Euclidean; standarisasi/normalisasi wajib dilakukan untuk mencegah bias.

2.8 Evaluasi Kualitas Clustering

Evaluasi kualitas clustering merupakan tahapan penting untuk menilai seberapa baik pola pengelompokan yang dihasilkan suatu algoritma. Berbeda dengan supervised learning yang memiliki label kelas sebagai acuan, penilaian kualitas cluster dilakukan melalui pendekatan internal validation, yaitu metode evaluasi yang mengukur kualitas hasil clustering berdasarkan struktur data internal tanpa memerlukan label. Evaluasi ini sangat penting karena menentukan apakah cluster yang terbentuk benar-benar mencerminkan karakteristik objek yang mirip, serta mendeteksi adanya pengelompokan yang kurang tepat. Dalam beberapa tahun terakhir, penelitian berfokus pada penggunaan metrik internal seperti Silhouette Coefficient, Davies-Bouldin Index, dan Dunn Index, namun Silhouette tetap menjadi indikator paling populer karena interpretasinya yang intuitif dan akurasinya yang stabil pada berbagai algoritma clustering (Arbelaitz et al., 2020; Javed et al., 2021).

2.8.1 Silhouette Coefficient

a) Konsep Dasar

Silhouette Coefficient atau Silhouette Score merupakan metrik yang digunakan untuk menilai seberapa tepat suatu titik berada pada cluster yang ditetapkan dibandingkan dengan kedekatannya pada cluster lain. Metrik ini mengombinasikan konsep compactness (kedekatan dalam cluster) dan separation (pemisahan antar cluster) sehingga mampu mengidentifikasi titik yang benar-benar terkelompok dengan baik, titik yang berada pada batas cluster, maupun titik yang salah klasifikasi. Metode ini pertama kali diperkenalkan oleh Rousseeuw (1987), namun penelitian modern membuktikan bahwa metrik ini tetap relevan dan unggul dalam mengevaluasi clustering pada data multidimensi maupun data berbasis Silhouette juga terbukti efektif digunakan pada berbagai algoritma seperti K-Means, DBSCAN, dan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hierarchical Clustering dalam studi-studi segmentasi industri dan UMKM

Rumus Silhouette Coefficient

Silhouette dihitung menggunakan persamaan dengan:

$$S(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \quad (2.9)$$

Keterangan:

- $a(i)$: rata-rata jarak antara titik data ke- i dengan seluruh titik data lain yang berada dalam klaster yang sama
- $b(i)$: jarak rata-rata terkecil antara titik data ke- i dengan klaster terdekat yang berbeda
- $S(i)$: nilai koefisien siluet untuk titik data ke- i

Nilai $S(i)$ berada pada rentang -1 hingga $+1$, di mana nilai mendekati $+1$ menunjukkan bahwa titik sangat tepat berada pada clusternya, nilai mendekati 0 menunjukkan bahwa titik berada pada batas dua cluster, sedangkan nilai negatif menunjukkan titik lebih dekat ke cluster lain sehingga mengindikasikan ketidaktepatan pengelompokan. Berbagai penelitian terbaru menunjukkan bahwa silhouette menjadi metrik internal paling stabil dalam kondisi data UMKM yang memiliki variabel campuran (omzet, usia usaha, tenaga kerja, kategori usaha) karena sifatnya yang tidak tergantung pada asumsi distribusi.

Interpretasi Nilai Silhouette

Nilai silhouette dapat digunakan untuk mengevaluasi kualitas cluster secara menyeluruh. Interpretasi umum yang digunakan dalam penelitian modern seperti pada Tabel 2.2 berikut:

Tabel 2.2. Interpretasi Nilai Silhouette

Nilai Silhouette	Interpretasi
0.71 – 1.00	Struktur <i>cluster</i> sangat baik
0.51 – 0.70	Struktur <i>cluster</i> kuat
0.26 – 0.50	Struktur <i>cluster</i> moderat/lemah
≤ 0.25	Struktur <i>cluster</i> buruk

Peneliti sering membandingkan nilai silhouette dari berbagai algoritma un-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengutip kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

tuk menentukan metode clustering terbaik. Misalnya, penelitian terbaru menunjukkan bahwa DBSCAN sering memperoleh nilai silhouette tinggi pada dataset dengan pola densitas, sedangkan K-Means unggul pada data berbentuk spherical

Kelebihan Silhouette Coefficient

Silhouette memiliki beberapa keunggulan, antara lain stabil pada berbagai jenis algoritma, tidak memerlukan label data, dan mampu memberikan gambaran global maupun lokal terkait kualitas pengelompokan. Metrik ini juga dapat divisualisasikan dalam bentuk diagram silhouette yang banyak digunakan dalam penelitian analisis segmentasi UMKM, data pemasaran, dan data sosial ekonomi karena mudah diinterpretasikan oleh analis maupun pemangku kebijakan

Kekurangan Silhouette Coefficient

Meski unggul, silhouette memiliki keterbatasan. Pada data berdimensi tinggi, jarak antar titik menjadi kurang representatif sehingga nilai silhouette dapat bias. Selain itu, silhouette kurang optimal untuk dataset yang memiliki cluster dengan bentuk sangat kompleks atau ukuran varians yang tidak seimbang, meskipun tetap dapat diaplikasikan pada DBSCAN dan hierarchical clustering. Sensitivitasnya terhadap pemilihan metrik jarak juga menjadi pertimbangan penting dalam analisis.

Relevansi Silhouette Untuk Penelitian UMKM

Dalam penelitian segmentasi UMKM Kota Pekanbaru, Silhouette Coefficient digunakan untuk memastikan bahwa cluster yang terbentuk benar-benar mencerminkan pola usaha perbedaan omzet, tingkat partisipasi ekatalog lokal, kapasitas tenaga kerja, dan usia usaha. Metrik ini juga berperan penting dalam membandingkan performa tiga algoritma yang digunakan (K-Means, DBSCAN, dan Hierarchical Clustering), memilih jumlah cluster optimal, dan memberikan dasar objektif dalam merumuskan rekomendasi kebijakan digitalisasi UMKM. Penggunaan silhouette dalam penelitian UMKM telah direkomendasikan oleh berbagai studi terbaru karena tingkat keandalannya dalam memvalidasi segmentasi berbasis data.

2.9 Orange

Orange merupakan perangkat lunak open-source yang dikembangkan sejak akhir 1990-an oleh Laboratory of Bioinformatics, Fakultas Ilmu Komputer dan Informasi, Universitas Ljubljana, Slovenia. Awalnya, Orange hanya berupa pustaka Python sederhana untuk keperluan analisis data di bidang bioinformatika. Seiring perkembangan kebutuhan analisis data, Orange kemudian berkembang menjadi sebuah toolbox data mining yang mendukung pemrosesan data, mesin pembelajaran, hingga visualisasi interaktif. Berikut adalah logo orange yang banyak di kenal Gambar 2.6



Gambar 2.6. Orange Data Mining

Tonggak penting terjadi pada tahun 2003, ketika Orange mulai menyediakan antarmuka berbasis visual programming. Hal ini membuat pengguna dapat merancang alur kerja analisis data (workflow) dengan metode drag-and-drop, sehingga tidak perlu lagi menulis kode pemrograman dari awal. Pada tahun 2013, Demšar dkk. menerbitkan artikel berjudul Orange: Data Mining Toolbox in Python di Journal of Machine Learning Research, yang memperkuat posisi Orange sebagai salah satu perangkat analisis data yang bersifat ilmiah sekaligus praktis (Demšar et al., 2013). Hingga kini, Orange terus dikembangkan dengan integrasi Python modern, pustaka scikit-learn, serta komunitas pengguna yang aktif.

Sebagai perangkat lunak analisis data, Orange memiliki berbagai fungsi utama yang menjadikannya relevan baik dalam penelitian akademik maupun industri:

1. *Data Preprocessing*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengujikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

menyediakan widget untuk pembersihan data (data cleaning), normalisasi, transformasi skala, penghapusan data kosong, hingga pemilihan fitur.

2. **Clustering** – mendukung berbagai algoritma pengelompokan, termasuk K-Means, Hierarchical Clustering, dan DBSCAN. Hasil clustering dapat ditampilkan dalam bentuk visual seperti scatter plot atau s
3. **Klasifikasi dan Regresi** – menyediakan pembelajaran terawasi algoritma seperti Decision Tree, Random Forest, Naïve Bayes, Logistic Regression, dan SVM.
4. **Evaluation** – memiliki metrik evaluasi seperti Silhouette Score, Davies-Bouldin Index, Cross-Validation, dan ROC Curve untuk menilai kinerja model.
5. **Visualisasi** – mempermudah pemahaman pola data dengan scatter plot, box plot, bar chart, heat map, hingga dendrogram interaktif.
6. **Integrasi Python** – meskipun Orange berbasis visual, pengguna tetap dapat menambahkan skrip Python untuk analisis yang lebih kompleks, sehingga fleksibel untuk pemula maupun peneliti berpengalaman.

Dalam penelitian ini, Orange dipilih karena menawarkan keseimbangan antara kemudahan penggunaan dan kelengkapan fungsi analisis data. Proses yang akan dilakukan mencakup:

- Mengimpor data UMKM dari Dinas Koperasi dan UMKM Pekanbaru, termasuk informasi mengenai kategori usaha, omzet, jumlah tenaga kerja, serta indikator partisipasi dalam e-katalog lokal.
- Melakukan preprocessing untuk membersihkan data, misalnya menghapus data duplikat, mengatasi data kosong, serta menormalkan atribut numerik seperti omzet.
- Mengaplikasikan algoritma K-Means Clustering melalui widget clustering untuk mengelompokkan UMKM menjadi beberapa segmen, seperti UMKM yang aktif di e-katalog, UMKM yang baru bergabung, dan UMKM yang belum berpartisipasi.
- Melakukan evaluasi kualitas hasil clustering menggunakan metrik seperti Silhouette Score atau Davies-Bouldin Index yang tersedia di Orange, sehingga dapat dipastikan bahwa segmentasi UMKM yang dihasilkan memiliki validitas.

- Menyajikan visualisasi hasil clustering dalam bentuk scatter plot atau siluet plot, sehingga hasil segmentasi lebih mudah dipahami oleh pemerintah daerah maupun pihak terkait.

Dengan fitur-fitur ini, Orange tidak hanya digunakan sebagai alat teknis untuk menjalankan algoritma, tetapi juga sebagai media interpretasi hasil penelitian. Hasil visualisasi yang dihasilkan Orange dapat memperjelas pola keterlibatan UMKM dalam e-katalog lokal, sehingga membantu pemerintah dalam menyusun strategi pelatihan yang lebih tepat sasaran.

Penggunaan Orange dalam akademi penelitian semakin banyak diadopsi karena sifatnya yang terbuka, interaktif, dan dapat diproduksi. Studi Demšar dkk. (2013) menunjukkan bahwa Orange dirancang untuk mendukung mesin pembelajaran dan eksplorasi data dalam konteks penelitian. Selain itu, dokumentasi resmi Orange (2024) menekankan bahwa perangkat ini tidak hanya bermanfaat bagi data scientist berpengalaman, tetapi juga untuk peneliti dari berbagai bidang yang membutuhkan analisis data berbasis alur kerja(Demsar dkk., t.t.).

2.10 Penelitian Terdahulu

Subbab ini menyajikan ringkasan penelitian terdahulu yang relevan dengan topik segmentasi dan digitalisasi UMKM, khususnya yang membahas pemanfaatan e-Katalog serta penerapan metode clustering seperti K-Means, DBSCAN, dan Hierarchical Clustering. Penelitian-penelitian tersebut yang telah di review pada Tabel 2.3 digunakan sebagai bahan pembanding untuk menunjukkan perbedaan pendekatan, metode analisis, serta kontribusi penelitian ini dalam menghasilkan segmentasi UMKM berbasis data yang mendukung perumusan kebijakan pembinaan.

Tabel 2.3. Penelitian Terdahulu Terkait UMKM, E-Katalog, dan Clustering

1	Judul Penelitian	Digitalisasi Produk UMKM Berbasis E-Katalog
	Peneliti & Tahun	(Irianto dkk., 2023)
	Metode	Studi kasus implementasi dengan wawancara pelaku UMKM
	Teori	Teori penerapan teknologi dan digitalisasi UMKM yang menekankan pentingnya literasi digital dan dukungan kelembagaan.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak mengikuti kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau	Hasil	E-katalog mempermudah promosi produk UMKM, namun rendahnya literasi digital, minim pelatihan teknis, dan keterbatasan infrastruktur menjadi hambatan utama.
	Judul Penelitian	Analisis Implementasi E-Katalog terhadap Perkembangan UMKM (Medan Marelan)
	Peneliti & Tahun	(Diani & Lubis, t.t.)
	Metode	Analisis deskriptif berbasis survei
3	Teori	Teori partisipasi UMKM dalam sistem e-procurement.
	Hasil	E-katalog memperluas akses pasar dan daya saing UMKM, namun partisipasi masih rendah karena keterbatasan pemahaman prosedur dan standar produk.
	Judul Penelitian	Implementasi Digitalisasi Produk UMKM dengan E-Katalog dan GIS
	Peneliti & Tahun	(Santoso dkk., 2022)
4	Metode	Sistem informasi terintegrasi (E-Katalog dan GIS)
	Teori	Teori sistem informasi manajemen dan pemetaan digital.
	Hasil	Integrasi e-katalog dan GIS memungkinkan pemetaan geografis produk UMKM untuk mendukung monitoring dan pengambilan keputusan.
	Judul Penelitian	MSME Clusterization Using K-Means in Garut
5	Peneliti & Tahun	(Alamanda dkk., 2023b)
	Metode	K-Means clustering
	Teori	Teori clustering dan segmentasi pasar berbasis data numerik.
	Hasil	UMKM terbagi ke dalam klaster besar, menengah, dan kecil berdasarkan omzet dan kapasitas usaha.
6	Judul Penelitian	Implementasi K-Means untuk Segmentasi UMKM
	Peneliti & Tahun	(Tony Hidayat, 2023)
	Metode	K-Means clustering
	Teori	Unsupervised learning dalam machine learning.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

- Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak Cipta milik UIN Suska Riau	Hasil	Terbentuk tiga klaster UMKM (aktif, berkembang, stagnan) sebagai dasar penentuan strategi pembinaan.
	Judul Penelitian	K-Means for Intelligent Customer Segmentation
	Peneliti & Tahun	(Tabianan dkk., 2022)
	Metode	K-Means clustering, Silhouette dan WCSS
	Teori	Teori perilaku konsumen dan evaluasi clustering.
	Hasil	K-Means membentuk segmen pelanggan dengan kualitas klaster yang baik.
	Judul Penelitian	Review of K-Means Algorithms
	Peneliti & Tahun	(Ikotun dkk., 2023)
	Metode	Review literatur
	Teori	Teori optimasi centroid dalam clustering.
8	Hasil	K-Means tetap populer meskipun sensitif terhadap inisialisasi centroid.
	Judul Penelitian	K-Means Clustering for Grouping Provinces (COVID-19)
	Peneliti & Tahun	(Abdullah dkk., 2022)
	Metode	K-Means clustering
	Teori	Teori analisis spasial dan data mining.
9	Hasil	Provinsi di Indonesia dikelompokkan berdasarkan indikator COVID-19.
	Judul Penelitian	Exploration of Clustering Algorithms for Customer Profiling
	Peneliti & Tahun	(John dkk., 2023)
	Metode	Perbandingan K-Means, Hierarchical, dan DBSCAN
	Teori	Teori unsupervised learning.
10	Hasil	K-Means paling efisien untuk dataset besar namun kurang optimal untuk bentuk klaster kompleks.
	Judul Penelitian	BI and Data Mining in SMEs
	Peneliti & Tahun	(Risco-Ramos dkk., 2023)
	Metode	Business Intelligence dan Data Mining
	Teori	Teori BI dan Knowledge Discovery in Databases.
11	Hasil	BI membantu UMKM mengubah data transaksi menjadi wawasan strategis.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

- b. Pengutipan tidak mengutip kepentingan yang wajar UIN Suska Riau.
- 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

⑩ Hak Cipta milik UIN Suska Riau	Judul Penelitian	Segmentasi Produk Toko UMKM Berbasis Restock
	Peneliti & Tahun	(Salsabiela dkk., 2024)
	Metode	K-Means clustering dan DBI
	Teori	Clustering untuk manajemen persediaan.
	Hasil	DBI = 0,436 menunjukkan kualitas klaster cukup baik.
	Judul Penelitian	Segmentasi Pelanggan Perusahaan Distribusi
⑪ Hak Cipta milik UIN Suska Riau	Peneliti & Tahun	(Mawarni dkk., 2023)
	Metode	K-Means clustering
	Teori	Teori segmentasi pelanggan dan CRM.
	Hasil	Pelanggan terbagi ke dalam segmen bernilai tinggi, sedang, dan rendah.
	Judul Penelitian	Segmentasi Pasar UMKM dengan K-Means
	Peneliti & Tahun	(Afrizal dkk., 2025)
⑫ Hak Cipta milik UIN Sultan Syarif Kasim Riau	Metode	K-Means, Silhouette, dan DBI
	Teori	Evaluasi clustering dan segmentasi pasar UMKM.
	Hasil	Silhouette 0,72 dan DBI 0,45 menunjukkan klaster valid dan stabil.
	Judul Penelitian	Segmentasi UMKM Pesisir Selatan Berdasarkan Sektor
	Peneliti & Tahun	(Arrahmi dkk., 2025)
	Metode	K-Means clustering dan Elbow Method
⑬ Hak Cipta milik UIN Sultan Syarif Kasim Riau	Teori	Teori segmentasi industri.
	Hasil	UMKM terbagi berdasarkan sektor usaha sebagai dasar kebijakan pembinaan.
	Judul Penelitian	Implementasi E-Katalog di Pekanbaru
	Peneliti & Tahun	(Ahmad, 2024a)
	Metode	Deskriptif kualitatif
	Teori	Teori adopsi teknologi e-procurement.
⑭ Hak Cipta milik UIN Sultan Syarif Kasim Riau	Hasil	E-katalog meningkatkan transparansi dan efisiensi, namun partisipasi UMKM masih rendah.
	Judul Penelitian	Studi Literatur Penyedia E-Katalog di LKPP
	Peneliti & Tahun	(Sinta Puspita Sari dkk., 2024)
	Metode	Studi literatur regulasi

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

- b. Pengutipan tidak mengikuti kepentingan yang wajar UIN Suska Riau.
- 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak Cipta milik UIN Suska Riau	Teori	Teori regulasi LKPP dan legalitas UMKM.
	Hasil	Mengidentifikasi persyaratan legal UMKM sebagai penyedia e-katalog.
	Judul Penelitian	Evaluation of K-Means, DBSCAN, and Hierarchical in SMEs
	Peneliti & Tahun	(A. Ramadhan dkk., 2025a)
	Metode	Perbandingan algoritma clustering
	Teori	Teori perbandingan algoritma dan validasi klaster.
© State Islamic University Sultan Syarif Kasim Riau	Hasil	K-Means unggul untuk data terstruktur, DBSCAN untuk noise, Hierarchical untuk relasi klaster.
	Judul Penelitian	Pengelompokan UMKM Pesisir Selatan dengan K-Means
	Peneliti & Tahun	(Tertiaavini, 2024)
	Metode	K-Means clustering dan Elbow
	Teori	Teori klasifikasi usaha UMKM.
	Hasil	UMKM terbagi berdasarkan jenis usaha, omzet, dan tenaga kerja.
© State Islamic University Sultan Syarif Kasim Riau	Judul Penelitian	K-Means Centroids Initialization Based on Differentiation
	Peneliti & Tahun	(Khan dkk., 2024a)
	Metode	Inisialisasi centroid
	Teori	Optimasi centroid dalam clustering.
	Hasil	Metode menghasilkan klaster lebih stabil dan akurat.
	Judul Penelitian	Research on the Evaluation of the Growth Level of SMEs Clustering Based on Digitalization
© State Islamic University Sultan Syarif Kasim Riau	Peneliti & Tahun	(Y. Wang, 2023)
	Metode	Fuzzy comprehensive evaluation dan clustering
	Teori	Teori digitalisasi UMKM dan evaluasi pertumbuhan.
	Hasil	Digitalisasi mempercepat pertumbuhan klaster SMEs dengan dukungan infrastruktur.
	Judul Penelitian	Impact of Digitalization on Technological Innovations in SMEs
	Peneliti & Tahun	(Radicic & Petković, 2023)
© State Islamic University Sultan Syarif Kasim Riau	Metode	Analisis kuantitatif panel data

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak meugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak Cipta milik UIN Suska Riau	Teori	Teori inovasi Schumpeterian dan transformasi digital.
	Hasil	Digitalisasi berpengaruh positif signifikan terhadap inovasi SMEs.
	Judul Penelitian	K-Means Centroids Initialization Based on Differentiation
	Peneliti & Tahun	(Khan dkk., 2024b)
	Metode	Perbandingan metode inisialisasi centroid
	Teori	Optimasi centroid clustering.
State Islamic University Sultan Syarif Kasim Riau	Hasil	Inisialisasi berbasis diferensiasi meningkatkan stabilitas klaster.
	Judul Penelitian	Adaptive K-Means Clustering Based Under-Sampling
	Peneliti & Tahun	(Zhou & Sun, 2024)
	Metode	K-Means adaptif
	Teori	Teori imbalance data dan evaluasi klaster.
	Hasil	Algoritma otomatis menentukan jumlah klaster pada data tidak seimbang.
24	Judul Penelitian	Comparative Study of K-Means-Based Techniques for SME Data
	Peneliti & Tahun	(Fareghi, 2025)
	Metode	Perbandingan varian K-Means
	Teori	Evaluasi clustering dan optimasi algoritma.
	Hasil	Terdapat trade-off antara kecepatan dan kualitas klaster.
	Judul Penelitian	k-LLMmeans: Scalable, Stable, and Interpretable Text Clustering
25	Peneliti & Tahun	(Diaz-Rodriguez, 2025)
	Metode	K-Means berbasis LLM
	Teori	Clustering modern dan interpretabilitas model.
	Hasil	Klaster teks lebih stabil dan interpretatif.
	Judul Penelitian	Application of the DBSCAN Algorithm in MSME Clustering
	Peneliti & Tahun	(A. Ramadhan dkk., 2025)
26	Metode	DBSCAN clustering

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak menghargai kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak Cipta milik UIN Suska Riau	Teori	Density-based clustering dan Silhouette.
	Hasil	DBSCAN efektif menangani outlier dan variasi kepadatan data MSME.
	Judul Penelitian	Evaluation of K-Means, DBSCAN, and Hierarchical Clustering
	Peneliti & Tahun	(Wikarta & Defiyanti, 2025)
	Metode	Perbandingan algoritma clustering
	Teori	Validasi klaster dan perbandingan algoritma.
© Statistic Islamic University Sultan Syarif Kasim Riau	Hasil	DBSCAN menghasilkan kualitas klaster terbaik, Hierarchical menengah, K-Means kurang optimal.
	Judul Penelitian	Pengelompokan UMKM Jawa Barat dengan K-Means
	Peneliti & Tahun	(Moehammad Nasri Abdoel Wahid dkk., 2024)
	Metode	K-Means clustering dan PCA
	Teori	Clustering dan distribusi wilayah.
	Hasil	UMKM terbagi menjadi tiga klaster dengan Silhouette sekitar 0,58.
© Statistic Islamic University Sultan Syarif Kasim Riau	Judul Penelitian	Analisis Klaster UMK Jawa Timur
	Peneliti & Tahun	(Wijaya dkk., 2025)
	Metode	Agglomerative Hierarchical Clustering
	Teori	Clustering hierarki dan segmentasi UMK.
	Hasil	Klaster membantu identifikasi kelompok UMK prioritas pembinaan.
	Judul Penelitian	Applied Machine Learning DBSCAN for Identifying Clusters
© Statistic Islamic University Sultan Syarif Kasim Riau	Peneliti & Tahun	(Sufairoh dkk., 2023)
	Metode	DBSCAN dan PCA
	Teori	Unsupervised learning dan reduksi dimensi.
	Hasil	DBSCAN efektif mengelompokkan industri mikro dan kecil.
	Judul Penelitian	Implementasi dan Perbandingan Clustering pada Data Marketplace
	Peneliti & Tahun	(Kurniadewi dkk., 2022)
© Statistic Islamic University Sultan Syarif Kasim Riau	Metode	K-Means, DBSCAN, dan Hierarchical

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

- Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau	Teori	Clustering dan segmentasi data marketplace.
	Hasil	Metodologi clustering marketplace relevan untuk data UMKM dan e-katalog.
	Judul Penelitian	Pemetaan UMKM untuk Pengentasan Kemiskinan
	Peneliti & Tahun	(Kurniadewi dkk., 2022)
	Metode	K-Means clustering
	Teori	Analisis sosial-ekonomi dan clustering UMKM.
	Hasil	Klaster UMKM membantu identifikasi wilayah prioritas intervensi.
	Judul Penelitian	Analisis Performa Clustering dengan Evaluasi Multi-Indeks
	Peneliti & Tahun	–

Berdasarkan sintesis terhadap 33 penelitian terdahulu, mayoritas penelitian segmentasi UMKM menggunakan algoritma *K-Means* dengan cakupan wilayah nasional maupun regional di Indonesia. Pendekatan ini umumnya diterapkan untuk mengelompokkan UMKM berdasarkan karakteristik sosial-ekonomi seperti omzet, jumlah tenaga kerja, dan skala usaha, serta terbukti menghasilkan struktur klaster yang relatif stabil pada *dataset* berukuran besar. Secara spasial, penelitian-penelitian tersebut tersebar pada wilayah nasional, daerah spesifik seperti Jawa Barat, Jawa Timur, Sidoarjo, dan Pekanbaru, serta sebagian kecil pada konteks internasional.

Selain *K-Means*, sejumlah penelitian juga menerapkan algoritma *DBSCAN* dan *Hierarchical Clustering* meskipun dalam jumlah yang lebih terbatas. Algoritma *DBSCAN* banyak digunakan pada data UMKM dengan kepadatan yang tidak merata dan keberadaan *outlier*, sehingga mampu membentuk klaster dengan bentuk yang tidak beraturan. Sementara itu, *Hierarchical Clustering* digunakan untuk menggambarkan hubungan hierarkis antar klaster UMKM, namun memiliki keterbatasan dari sisi efisiensi komputasi pada data berskala besar. Beberapa penelitian juga melakukan perbandingan lebih dari satu algoritma *clustering* untuk mengevalu-

uasi performa dan kualitas klaster yang dihasilkan.

Di sisi lain, penelitian yang berfokus pada digitalisasi UMKM dan e-katalog umumnya menggunakan pendekatan deskriptif dan studi kasus. Hasil penelitian menunjukkan bahwa e-katalog berkontribusi positif terhadap peningkatan akses pasar, transparansi transaksi, dan efisiensi proses pengadaan. Namun demikian, rendahnya literasi digital, keterbatasan pelatihan teknis, serta kesiapan infrastruktur masih menjadi kendala utama bagi partisipasi UMKM secara optimal.

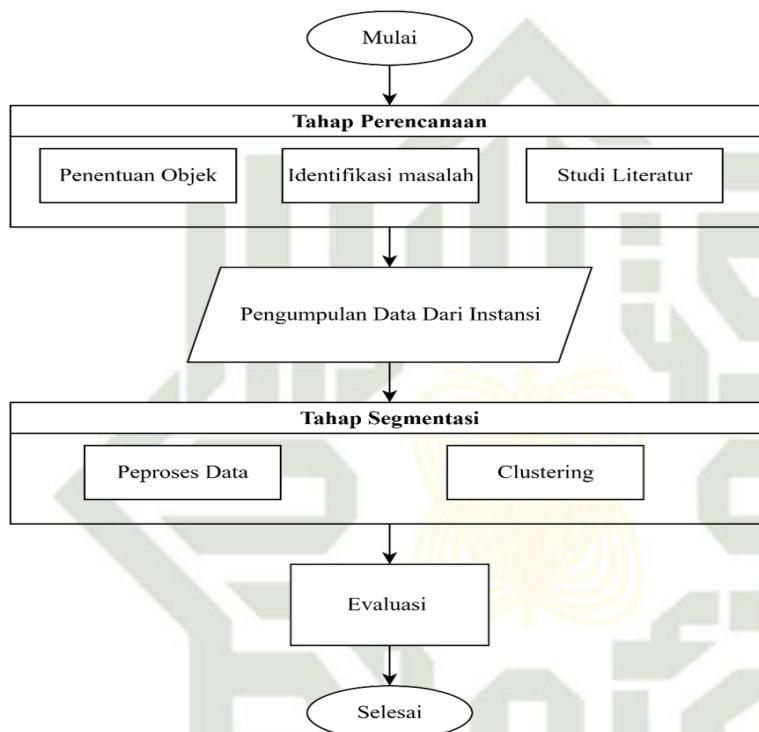
Meskipun demikian, sangat sedikit penelitian yang mengintegrasikan analisis partisipasi UMKM dalam e-katalog dengan segmentasi UMKM berbasis perbandingan multi-algoritma *clustering*, khususnya pada tingkat kota. Oleh karena itu, penelitian ini memiliki kebaruan dengan mengombinasikan analisis partisipasi UMKM dalam e-katalog lokal dan segmentasi menggunakan algoritma *K-Means*, *DBSCAN*, dan *Hierarchical Clustering* pada konteks UMKM Kota Pekanbaru, sehingga menghasilkan pemetaan yang lebih komprehensif sebagai dasar perumusan kebijakan pembinaan UMKM berbasis data.

BAB 3

METODOLOGI PENELITIAN

3.1 Alur Penelitian

Alur penelitian menggambarkan tahapan-tahapan sistematis yang dilakukan dalam penelitian ini mulai dari perencanaan hingga analisis hasil. Berikut adalah alur penelitian yang berada dalam Gambar 3.1:



Gambar 3.1. Alur Penelitian

3.1.1 Tahap Perencanaan

Tahap perencanaan merupakan langkah awal dalam penelitian yang bertujuan untuk menetapkan arah dan ruang lingkup penelitian. Pada tahap ini terdapat tiga kegiatan utama, yaitu penentuan objek penelitian, identifikasi masalah, dan studi literatur.

a) Penentuan Objek

Objek penelitian ini adalah UMKM di Kota Pekanbaru yang terdaftar pada Dinas koperasi usaha kecil dan menengah Kota Pekanbaru. Pemilihan objek ini didasarkan pada peran penting UMKM dalam mendorong perekonomian daerah, serta adanya kebijakan pemerintah mengenai penerapan e-katalog

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengujikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

lokal untuk meningkatkan partisipasi UMKM dalam pengadaan barang dan jasa (Ahmad, 2024b)

b) IdentifikasiMasalah

Proses identifikasi masalah pada umumnya mencakup tiga tahap, yaitu: menemukan permasalahan yang ada (problem), menelusuri akar penyebabnya (root cause), serta merumuskan pernyataan masalah (problem statement) yang menggambarkan secara jelas permasalahan yang telah diidentifikasi. Pada penelitian ini identifikasi masalah dilakukan untuk menemukan isu utama yang dihadapi UMKM, khususnya terkait rendah atau tidaknya tingkat partisipasi dalam e-katalog lokal. Tahap ini dilakukan dengan mengamati kondisi eksisting serta meninjau data awal yang tersedia, sehingga dapat dirumuskan permasalahan yang relevan sebagai dasar penelitian.

c) Studi Literatur

Studi literatur dilakukan dengan meninjau teori, penelitian terdahulu, serta regulasi yang relevan dengan topik penelitian. Kajian ini mencakup konsep dasar UMKM sebagai objek penelitian (Harahap et al., 2025), regulasi terkait digitalisasi layanan pemerintah melalui e-katalog lokal (Ahmad, 2024b), serta teori mengenai data mining dan algoritma K-Means Clustering sebagai metode analisis. Tahap ini bertujuan untuk memperkuat lansasan teoritis penelitian dan memperjelas posisi penelitian dalam konteks akademik maupun praktis.

3.1.2 Tahap Pengumpulan Data

Tahap pengumpulan data dalam penelitian ini dilakukan dengan menggunakan data sekunder yang diperoleh dari dua instansi pemerintah, yaitu Dinas UKM Kota Pekanbaru serta LPSE.

Data dari Dinas Koperasi, UKM Kota Pekanbaru digunakan untuk memperoleh informasi mengenai profil Usaha Mikro, Kecil, dan Menengah (UMKM) yang telah terverifikasi di wilayah Kota Pekanbaru. Data tersebut bersifat tabular dan diperoleh dalam format spreadsheet (.xlsx/.csv). Berdasarkan hasil pengumpulan, total terdapat 10.560 data UMKM terverifikasi pada tahun 2024.

Sementara itu, data pendukung lainnya diperoleh dari LPSE (Layanan Pengadaan Secara Elektronik) yang memuat data e-Katalog dan penyedia barang/jasa (Barjas). Data ini digunakan untuk mengidentifikasi keterlibatan UMKM dalam sistem katalog elektronik pemerintah, khususnya dalam konteks pengadaan barang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengikuti kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dan jasa publik. Kedua sumber data tersebut saling melengkapi, di mana data dari Dinas Koperasi digunakan untuk menggambarkan populasi dan karakteristik UMKM di Kota Pekanbaru, sedangkan data dari LPSE digunakan untuk menganalisis partisipasi UMKM dalam ekosistem digital pengadaan barang/jasa melalui e-Katalog. Dengan demikian, penggunaan data sekunder ini memungkinkan untuk melakukan analisis yang komprehensif terhadap pola partisipasi dan keterlibatan UMKM lokal dalam sistem ekonomi digital berbasis katalog pemerintah. Adapun variabel-variabel utama yang digunakan dalam penelitian ini antara lain:

1. **Nama UMKM** (Identitas unit usaha (tidak digunakan dalam proses clustering, hanya sebagai penanda)).
2. **Kategori Usaha** (Jenis bidang usaha (misalnya kuliner, jasa, kerajinan, perdagangan, dll.).
3. **Jumlah Tenaga Kerja** (Jumlah pekerja dalam UMKM).
4. **Omzet** (Rata-rata omzet per tahun).
5. **Lama Berdiri** (Usia usaha dalam tahun).
6. **Partisipasi dalam e-Katalog Lokal** (Status keikutsertaan UMKM dalam sistem e-katalog (misalnya: sudah terdaftar / belum terdaftar)).

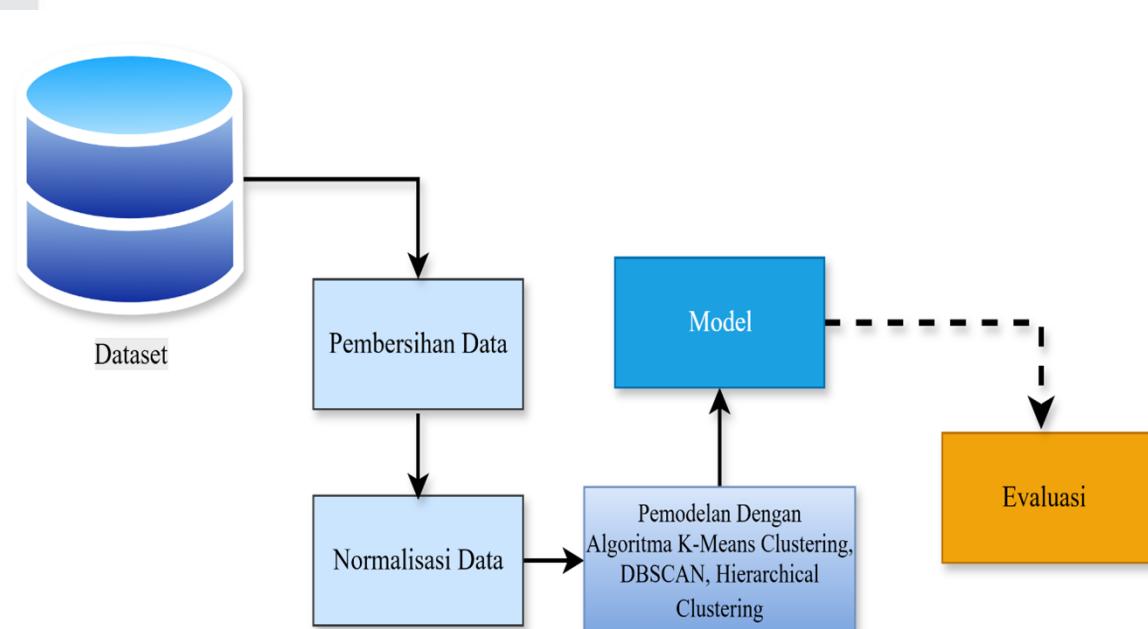
Dari variabel-variabel tersebut, hanya variabel numerik (jumlah tenaga kerja, omzet, lama berdiri) dan variabel kategori terukur (kategori usaha, status e-katalog) yang diproses lebih lanjut pada tahap segmentasi menggunakan algoritma K-Means Clustering. Variabel identitas seperti nama UMKM hanya berfungsi sebagai penanda hasil, bukan bagian dari proses analisis.

3.1.3 Tahap Segmentasi

Pada tahap ini dilakukan proses pengelompokan data menggunakan tiga algoritma, yaitu K-Means, DBSCAN, dan Hierarchical Clustering yang alur segmentasinya dapat di lihat dalam Gambar 3.2. Pemodelan dilaksanakan melalui Orange Data Mining, yang digunakan untuk melakukan normalisasi data, menjalankan ketiga algoritma, serta mengevaluasi hasil klaster menggunakan Silhouette Score dan visualisasi scatter plot atau dendrogram. Tahap ini bertujuan membentuk segmentasi UMKM berdasarkan kemiripan karakteristik usaha dan tingkat partisipasi dalam e-katalog lokal. Melalui penerapan ketiga algoritma tersebut, penelitian memperoleh gambaran klaster yang lebih komprehensif dan dapat membandingkan pola pengelompokan dari pendekatan centroid-based, density-based, dan hierarchical:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengutip kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 3.2. Tahap *Clustering*

3.1.3.1 Pra Pemrosesan Data

Pra-pemrosesan data merupakan tahap penting dalam penelitian ini yang bertujuan untuk menyiapkan dataset agar siap digunakan dalam proses clustering. Tahap ini mencakup serangkaian langkah untuk meningkatkan kualitas data, mengurangi kesalahan, serta memastikan konsistensi antarvariabel. Dengan melakukan pra-pemrosesan, data menjadi lebih bersih, terstruktur, dan sesuai dengan kebutuhan algoritma K-Means Clustering, sehingga hasil analisis yang diperoleh lebih akurat, valid, dan dapat diandalkan (Guo dkk., 2023; Côté dkk., 2024)

a) Pembersihan Data

Pembersihan data dilakukan dengan cara memeriksa adanya nilai yang hilang (missing values), data ganda (duplicate records), maupun data yang tidak konsisten dengan variabel penelitian. Data yang tidak lengkap diperbaiki melalui teknik imputasi sederhana atau dihapus apabila tidak relevan, sedangkan data ganda dieliminasi agar tidak menimbulkan bias. Tahap ini memastikan bahwa dataset yang digunakan hanya berisi data yang valid dan representatif untuk dianalisis.

b) Normalisasi Data

Setelah pembersihan dilakukan, tahap selanjutnya adalah normalisasi. Normalisasi diperlukan untuk menyeragamkan skala antarvariabel agar tidak ada variabel yang mendominasi perhitungan jarak dalam algoritma K-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengurangi kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Means Clustering. Pada penelitian ini digunakan teknik Min-Max Normalization yang mengubah nilai data ke dalam rentang [0,1]. Langkah ini penting karena kualitas hasil clustering sangat dipengaruhi oleh skala fitur yang digunakan (Wongoutong, 2024; Maharana et al., 2022).

3.1.3.2 Pemodelan Dengan Algoritma K-Means Clustering

Tahap pemodelan merupakan inti dari penelitian ini, yaitu penerapan algoritma K-Means Clustering untuk mengelompokkan data UMKM di Kota Pekanbaru berdasarkan karakteristik tertentu. Algoritma ini dipilih karena relatif sederhana, efisien, dan terbukti banyak digunakan dalam penelitian serupa untuk tujuan segmentasi data (Patel, Prajapati, 2018; Jain, 2010; Ullmann et al., 2021). Pemodelan menggunakan algoritma K-Means dilakukan dengan bantuan aplikasi Orange Data Mining. K-Means dipilih sebagai algoritma utama karena memiliki kinerja yang cepat, efisien, serta banyak digunakan dalam penelitian segmentasi UMKM dan data skala besar (Patel, Prajapati, 2018; Jain, 2010; Ullmann et al., 2021). Dalam penelitian ini, proses pemodelan K-Means dilakukan melalui langkah-langkah berikut:

1. Menentukan Nilai K (Jumlah Cluster)

Penelitian tidak langsung menetapkan jumlah cluster, melainkan melakukan beberapa percobaan nilai k (misalnya $k = 2$ hingga $k = 5$). Setiap nilai k kemudian dievaluasi menggunakan Silhouette Score pada Orange untuk menentukan jumlah cluster yang paling optimal.

2. Mengatur Parameter K-Means pada Orange

Orange secara otomatis melakukan inisialisasi centroid dan proses iterasi. Peneliti hanya mengatur parameter seperti:

- jumlah cluster (k),
- maksimum iterasi,
- metode inisialisasi centroid (default Orange).

Dengan demikian, seluruh proses komputasi centroid dan pembaruan jarak ditangani oleh sistem.

3. Menjalankan Proses Clustering

Orange menghitung jarak antar data menggunakan Euclidean Distance dan mengelompokkan UMKM ke dalam klaster berdasarkan kemiripan karakteristik (omzet, lama usaha, partisipasi e-katalog, jumlah tenaga kerja, kategori usaha).

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengutip kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. **Mengenali Pola Klaster yang Terbentuk** Hasil pemodelan ditampilkan dalam bentuk

- daftar label klaster,
- Scatter Plot,
- Silhouette Plot, yang memudahkan

peneliti memahami karakteristik tiap klaster.

Menentukan Cluster Terbaik

Nilai Silhouette Score dari setiap percobaan k digunakan untuk memilih jumlah klaster yang paling representatif. Nilai $\geq 0,5$ dianggap valid dan menunjukkan pemisahan klaster yang baik. Dengan pemodelan melalui Orange, proses clustering dapat dilakukan secara konsisten tanpa perhitungan manual, serta memungkinkan visualisasi yang jelas terhadap hasil segmentasi UMKM.

3.1.3.3 Pemodelan Dengan Algoritma Dbscan

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) digunakan sebagai algoritma kedua karena mampu membentuk klaster berdasarkan kepadatan data dan dapat mendeteksi noise atau objek yang tidak termasuk dalam klaster mana pun. Algoritma ini relevan untuk data UMKM yang kemungkinan memiliki distribusi tidak homogen (Ester dkk., 1996b) Pemodelan DBSCAN dilakukan menggunakan Orange Data Mining dengan langkah berikut:

1. **Menentukan Parameter Utama**

Pada Orange, dua parameter utama DBSCAN yang harus ditetapkan adalah:

- eps (radius pencarian sekitar tiap titik)
- min samples (jumlah minimum titik di sekitar radius untuk membentuk klaster).

Penelitian ini melakukan beberapa percobaan kombinasi parameter untuk menemukan konfigurasi yang menghasilkan klaster yang paling stabil.

2. **Menjalankan Proses Clustering**

Orange akan mengelompokkan UMKM berdasarkan kepadatan. Titik-titik yang tidak memiliki tetangga yang cukup akan dianggap sebagai noise. Hal ini memungkinkan analisis tambahan terkait UMKM yang memiliki karakteristik sangat berbeda atau belum siap berpartisipasi dalam sistem digital seperti e-katalog.

3. **Visualisasi Hasil Klaster** Hasil pemodelan divisualisasikan menggunakan:

- Scatter Plot (untuk melihat bentuk klaster),
- Silhouette Plot (untuk mengevaluasi kualitas klaster),
- informasi jumlah klaster dan noise.

4. **Menentukan Model Terbaik**

Model DBSCAN yang baik dipilih berdasarkan nilai Silhouette Score tertinggi dari percobaan beberapa parameter. Visualisasi Scatter Plot juga digunakan untuk memastikan klaster tidak terlalu kecil atau terlalu terfragmentasi.

Melalui DBSCAN, penelitian dapat mengidentifikasi klaster UMKM berbasis kepadatan serta mengamati kelompok UMKM yang berpotensi menjadi outlier atau memiliki kondisi usaha sangat berbeda.

3.1.3.4 Pemodelan Dengan Algoritma Hierarchical Clustering

Hierarchical Clustering digunakan sebagai algoritma ketiga karena mampu menunjukkan hubungan antar data secara bertingkat (hierarki). Metode ini memberikan gambaran lebih rinci tentang bagaimana UMKM saling terhubung pada berbagai tingkat pengelompokan. Pemodelan Hierarchical Clustering dilakukan melalui Orange dengan langkah berikut:

1. **Memilih Metode Penggabungan (Linkage)**

Dalam aplikasi Orange, analisis hierarchical clustering dapat dilakukan dengan berbagai kriteria linkage seperti single linkage, complete linkage, dan average linkage, yang masing-masing memiliki ciri khas dalam menentukan jarak antar klaster. Single linkage menilai kedekatan dua klaster berdasarkan jarak terpendek antar pasangan objek, sehingga sering menyebabkan efek chaining, yaitu terbentuknya klaster memanjang akibat penggabungan objek-objek yang berdekatan meskipun secara keseluruhan cukup berjauhan. Sebaliknya, complete linkage menggunakan jarak terjauh antar anggota klaster sebagai dasar penggabungan, sehingga biasanya menghasilkan klaster yang lebih kompak dan terpisah dengan jelas. Sementara itu, average linkage menggunakan rata-rata jarak seluruh pasangan objek dari dua klaster sebagai ukuran kedekatan, sehingga memberikan kestabilan antara kecenderungan chaining dan kekompakkan klaster yang dihasilkan (Wicaksana, Adikara, dan Adinugroho, 2018).

Pada penelitian ini, berbagai kriteria linkage tersebut dianalisis untuk mengamati struktur klaster yang terbentuk dan memilih konfigurasi yang

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengurangi kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

paling merepresentasikan karakteristik data UMKM. Evaluasi dilakukan secara visual melalui interpretasi dendrogram dan analisis distribusi variabel pada tiap klaster, dengan penekanan pada homogenitas internal serta keterpisahan antar kelompok. Berdasarkan hasil evaluasi tersebut, Ward's linkage dipilih sebagai metode yang digunakan dalam analisis akhir. Ward's linkage menerapkan prinsip minimum variance dengan meminimalkan kelebihan jumlah kuadrat galat (within-cluster sum of squares) pada setiap tahap penggabungan, sehingga cenderung menghasilkan klaster yang lebih homogen dan seimbang dalam jumlah anggota(Ishanifa, 2025)

Ward's linkage dianggap paling cocok untuk data numerik yang telah di-normalisasi karena menggunakan jarak Euclidean kuadrat dan memiliki karakteristik yang serupa dengan tujuan pada metode k-means, meskipun diimplementasikan dalam kerangka hierarchical *agglomerative clustering*. Hal ini membuat pemisahan klaster menjadi lebih stabil, struktur dendrogram lebih teratur, serta hasil segmentasi UMKM lebih mudah diinterpretasikan untuk mendukung perumusan strategi pembinaan maupun pengambilan keputusan berbasis data(Madhulatha, 2021)

2. Membangun Struktur Hirarki

Orange menghasilkan dendrogram yang memperlihatkan proses penggabungan klaster dari level paling kecil hingga besar. Dendrogram menjadi alat utama untuk memahami karakteristik data, termasuk kelompok UMKM yang memiliki kedekatan atribut tertentu.

3. Menentukan Jumlah Klaster

Pemotongan dendrogram dilakukan untuk menentukan jumlah klaster optimal. Hasil pemotongan ini menghasilkan label klaster untuk setiap data UMKM.

4. Evaluasi Menggunakan Silhouette Score

Setiap hasil pemotongan klaster dievaluasi dengan Silhouette Score untuk menilai kualitas pemisahan klaster. Nilai di atas 0,5 menunjukkan pengelompokan yang baik dan digunakan sebagai acuan dalam menentukan konfigurasi final.

5. Visualisasi Klaster

Hasil akhir divisualisasikan menggunakan Scatter Plot untuk melihat pemisahan antar klaster setelah dendrogram dipotong menjadi klaster final. Hierarchical Clustering memberikan wawasan tambahan mengenai kedekatan antar data UMKM yang mungkin tidak terlihat pada algoritma

berbasis centroid seperti K-Means.

3.1.4 Evaluasi

Tahap evaluasi dilakukan untuk menilai sejauh mana hasil clustering yang diperoleh mampu menggambarkan segmentasi UMKM secara tepat. Evaluasi ini bertujuan untuk memastikan bahwa cluster yang terbentuk benar-benar mencerminkan kondisi partisipasi UMKM dalam e-katalog lokal, sehingga dapat dijadikan dasar analisis dan penarikan kesimpulan penelitian.

Evaluasi hasil clustering dalam penelitian ini dilakukan menggunakan Silhouette Score digunakan untuk menilai kualitas hasil clustering, dengan rentang nilai -1 hingga 1. Nilai mendekati 1 menunjukkan bahwa cluster terbentuk dengan baik, sedangkan nilai negatif menandakan adanya kesalahan dalam pengelompokan.

Dalam penelitian ini, cluster dinyatakan valid apabila memiliki nilai Silhouette $\geq 0,5$, yang berarti anggota dalam cluster cukup homogen dan berbeda jelas dengan cluster lain. Dengan indikator tersebut, hasil clustering tidak hanya diuji secara kuantitatif tetapi juga dapat diinterpretasikan secara kontekstual untuk memberikan gambaran segmentasi UMKM yang lebih akurat.

Hasil evaluasi diharapkan dapat memberikan informasi yang valid mengenai jumlah cluster yang paling sesuai dengan data, kualitas pemisahan antar cluster, serta keseragaman anggota di dalam setiap cluster. Dengan demikian, evaluasi berperan penting dalam menjamin keandalan hasil segmentasi sebelum dilakukan analisis lebih lanjut dan penyusunan rekomendasi.

3.2 Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan deskriptif. Penelitian kuantitatif digunakan karena data yang dianalisis bersifat numerik dan dapat diolah secara statistic, Penelitian kuantitatif merupakan metode ilmiah yang memanfaatkan data berbentuk angka sebagai dasar dalam proses pengumpulan dan analisis. Melalui data numerik tersebut, penelitian ini dapat digunakan untuk menguji hipotesis, menganalisis hubungan antar variabel, membuat prediksi, serta melakukan generalisasi temuan ke populasi yang lebih luas dengan dukungan teknik statistik dan instrumen terstruktur. Pendekatan ini juga menitik beratkan pada objektivitas, pengukuran yang tepat, serta pengendalian variable (Jauza dan Albina, 2025).

Sedangkan Pendekatan deskriptif dipilih dalam penelitian ini karena bertujuan memberikan gambaran sistematis, faktual, dan akurat mengenai kondisi partisipasi UMKM di e-katalog lokal. Menggunakan pendekatan ini memungkinkan

- Hak Cipta Bilindungj Undang-Undang**

 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengutip kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

kan peneliti untuk fokus pada pengumpulan data numerik dan observasi situasi sekarang, tanpa mencoba menentukan hubungan sebab-akibat. Dengan demikian, data yang diperoleh akan memperlihatkan apa yang terjadi, bagaimana keadaan saat ini, dan berapa banyak UMKM yang berpartisipasi, sebagai dasar bagi rekomendasi kebijakan dan penelitian selanjutnya.

Penelitian ini juga termasuk dalam kategori data mining eksploratif, karena bertujuan untuk menemukan pola atau pengelompokan dari data yang tersedia menggunakan algoritma K-Means Clustering. Hasil segmentasi ini diharapkan dapat memberikan informasi yang berguna dalam pengambilan keputusan, khususnya dalam meningkatkan partisipasi UMKM dalam digitalisasi layanan pemerintah.

Penelitian ini menggunakan algoritma K-Means Clustering untuk melakukan segmentasi data UMKM berdasarkan variabel numerik yang diperoleh dari dataset. Metode ini dipilih karena mampu mengelompokkan data berdasarkan kesamaan karakteristik, sehingga cocok digunakan untuk analisis tingkat partisipasi UMKM dalam e-katalog lokal.

3.3 Alat dan Perangkat Lunak

Penelitian ini menggunakan beberapa perangkat lunak pendukung untuk proses pengolahan data, pemodelan, dan visualisasi hasil, yaitu:

a) **Orange Data Mining 3**

Orange digunakan sebagai platform utama dalam proses analisis data, karena menyediakan antarmuka visual berbasis workflow yang memudahkan implementasi data mining tanpa coding. Dalam penelitian ini, Orange digunakan untuk:

- reprocessing data (Select Columns, Impute, Normalize, Edit Domain)
- Visualisasi awal (Distribution, Box Plot, Heatmap)
- Penerapan algoritma clustering K-Means melalui widget k-Means, DBSCAN melalui widget DBSCAN Clustering, Hierarchical Clustering melalui widget Hierarchical Clustering
- Evaluasi model menggunakan Silhouette Score
- Visualisasi hasil klaster melalui Scatter Plot, Silhouette Plot, dan other multivariate tools.

b) **Spreadsheet**

Digunakan untuk: Pembersihan awal dataset, menyimpan file dalam format CSV sebelum diimpor ke Orange, perhitungan tambahan seperti pendapatan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengurangi kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

per tenaga kerja (omzet ÷ jumlah tenaga kerja)

Bahan Penelitian

Dataset UMKM tahun 2024 yang diperoleh dari Dinas Koperasi dan UMKM Kota Pekanbaru. Data partisipasi e-katalog lokal yang diperoleh dari LPSE Pekanbaru. Data mencakup variabel omzet, jumlah tenaga kerja, lama berdiri, kategori usaha, dan status partisipasi e-katalog.

3.4 Lokasi dan Waktu Penelitian

Lokasi penelitian ini dilakukan pada Dinas Koperasi UMK Kota Pekanbaru serta LPSE Kota Pekanbaru sebagai sumber data sekunder tahun 2024. Seluruh proses analisis dan pemodelan dilakukan secara mandiri menggunakan perangkat lunak Orange Data Mining.

Waktu penelitian dilaksanakan mulai 7 November 2025 (setelah peneliti memperoleh persetujuan pada seminar proposal) hingga Desember 2025, mencakup tahap pengumpulan data, preprocessing, pemodelan clustering (K-Means, DBSCAN, dan Hierarchical Clustering), evaluasi model, serta penyusunan laporan akhir.

BAB 5

PENUTUP

5.1 Kesimpulan

Penelitian ini berhasil menerapkan serta membandingkan algoritma *K-Means*, *DBSCAN*, dan *Hierarchical Clustering* dalam melakukan segmentasi UMKM di Kota Pekanbaru berdasarkan karakteristik usaha dan tingkat partisipasi dalam e-Katalog Lokal. Hasil segmentasi menunjukkan adanya perbedaan karakteristik yang jelas antar klaster, baik dari sisi kapasitas ekonomi, kesiapan digital, maupun tingkat keterlibatan dalam pengadaan pemerintah. Dengan demikian, pendekatan *clustering* terbukti mampu merepresentasikan kondisi UMKM secara objektif, sistematis, dan terstruktur.

Evaluasi kualitas klaster menggunakan *Silhouette Score* menunjukkan bahwa algoritma *K-Means* menghasilkan struktur klaster yang paling stabil dan mudah diinterpretasikan, dengan nilai *Silhouette Score* sebesar 0,444. Sementara itu, algoritma *DBSCAN* memperoleh nilai *Silhouette Score* sebesar 0,666, dan *Hierarchical Clustering* mencapai nilai tertinggi sebesar 0,693 pada konfigurasi klaster terbaik. Meskipun nilai evaluasi *Hierarchical Clustering* dan *DBSCAN* relatif lebih tinggi, algoritma *K-Means* dinilai lebih konsisten dan aplikatif untuk kebutuhan perumusan kebijakan publik karena mampu menghasilkan pembagian klaster yang jelas, seimbang, dan mudah dianalisis oleh pemangku kepentingan.

Hasil segmentasi UMKM yang diperoleh dalam penelitian ini dapat dijadikan dasar bagi pemerintah daerah dalam merancang strategi pembinaan UMKM yang lebih tepat sasaran dan berbasis data. Temuan penelitian menunjukkan adanya hubungan positif antara kapasitas usaha dan tingkat partisipasi dalam e-Katalog Lokal. Oleh karena itu, UMKM dengan skala usaha yang lebih kecil memerlukan pendampingan yang lebih intensif, khususnya dalam peningkatan literasi digital dan pemanfaatan platform pengadaan elektronik, guna mendorong transformasi digital serta meningkatkan daya saing UMKM secara berkelanjutan.

5.2 Saran

Pemerintah Kota Pekanbaru disarankan untuk menerapkan program pembinaan UMKM berbasis klaster sebagai strategi pengembangan yang lebih terarah dan efektif. UMKM pada *Cluster C1* memerlukan peningkatan literasi digital serta pendampingan intensif terkait pemanfaatan e-Katalog Lokal agar mampu berpartisipasi secara aktif dalam sistem pengadaan pemerintah. Selanjutnya, UMKM pada *Clus-*

ter C2 membutuhkan dukungan lanjutan dalam peningkatan kualitas produk, peningkatan pemasaran digital, serta diversifikasi pasar guna meningkatkan daya saing usaha. Sementara itu, UMKM pada *Cluster C3* memerlukan program revitalisasi yang mengintegrasikan digitalisasi usaha dengan pengembangan produk dan perluasan akses pasar secara berkelanjutan.

Selain itu, pemerintah daerah juga disarankan untuk memperkuat infrastruktur digital pendukung, menyediakan pusat pelatihan teknologi yang mudah diakses oleh UMKM, serta mengembangkan sistem pemantauan berbasis data guna mengevaluasi efektivitas program pembinaan dan tingkat partisipasi UMKM dalam e-Katalog Lokal secara berkelanjutan.

DAFTAR PUSTAKA

- Adian, I., Doumbia, D., Gregory, N., Ragoussis, A., Reddy, A., dan Timmis, J. D. (2020). *Small and medium enterprises in the pandemic: impact, responses and the role of development finance* (Tech. Rep.). The World Bank.
- Afrizal, M., Saputra, I., dan Satria, R. (2025). Analisis performa algoritma k-means clustering untuk segmentasi pasar di umkm. , 5(2). (Metadata jurnal belum dicantumkan secara lengkap)
- Ahmad, A. (2024). Implementasi e katalog terhadap perkembangan umkm di kota pekanbaru.
- Ahmed, M., Seraj, R., dan Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8), 1295.
- Alamanda, D. T., Kusmiati, E., Shiddieq, D. F., dan Roji, F. F. (2023). Msme clusterization using k-means clustering in garut regency, indonesia.
- Alzami, F., Sambasri, F. D., Kiswanto, R. M., Megantara, R. A., Akrom, A., Pramunendar, R. A., ... Sulistiyawati, P. (2022). Implementation of etl e-commerce for customer clustering using rfm and k-means clustering. *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)*, 10(3), 167.
- Andira, A. P., dan Chandra, R. (2023). Pengaruh pertumbuhan usaha mikro kecil dan menengah (umkm) terhadap pertumbuhan ekonomi kota pekanbaru tahun 2020–2022.
- Azzam, A., Purnamasari, A. I., dan Ali, I. (2024). Implementasi algoritma k-means clustering untuk analisis persebaran umkm di jawa barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3062–3070.
- Burdiono, W., Nugroho, B. I., dan Santoso, N. A. (2024). Penerapan metode fuzzy k-means clustering untuk pengelompokan konten halaman web secara otomatis.
- Ester, M., Kriegel, H.-P., dan Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise.
- Fahlevi, M. R., Putri, D. R. D., dan Syahrin, E. (2023). Analisis pengelompokan data pelelangan barang dengan metode k-means clustering.
- Faiz, F., Le, V., dan Masli, E. K. (2024). Determinants of digital technology adoption in innovative smes. *Journal of Innovation & Knowledge*, 9(4), 100610.
- Fikri, M. A. (2024). Review paper data mining metode clustering menggunakan algoritma k-means. *Jurnal Informatika*, 1(2).

- Husna, I. N. (2020). Kebijakan peningkatan usaha mikro, kecil, dan menengah di indonesia. *Muttaqien: Indonesian Journal of Multidiciplinary Islamic Studies*, 1(1), 43–55.
- Iin, I., Fadila, R., Rinaldi, A. R., dan Fathurrohman, F. (2024). Penerapan data mining dalam mengelompokan jumlah umkm berdasarkan kabupaten kota menggunakan k-means clustering. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1446–1450.
- Ishanifa, M. (2025). Use of hierarchical clustering method with complexity invariant distance on provincial rice prices in indonesia. *Journal of Applied Statistics and Data Science*, 2(1), 45–57.
- Jauza, N. A., dan Albina, M. (2025). Model dan pendekatan penelitian kuantitatif: Kajian filosofis, metodologis, dan aplikatif.
- Jelita, T., Buaton, R., dan Simanjuntak, M. (2023). Pengelompokan bidang usaha terhadap bantuan produktif usaha mikro (bpum) berdasarkan wilayah deli serdang menggunakan metode clustering k-means (studi kasus: Dinas koperasi dan umkm kabupaten deli serdang). *Explorer*, 3(2), 50–57.
- John, J. M., Shobayo, O., dan Ogunleye, B. (2023). An exploration of clustering algorithms for customer segmentation in the uk retail market. *Analytics*, 2(4), 809–823.
- Kahveci, E. (2025). Digital transformation in smes: Enablers, interconnections, and a framework for sustainable competitive advantage. *Administrative Sciences*, 15(3), 107.
- Khan, A. A., Bashir, M. S., Batool, A., Raza, M. S., dan Bashir, M. A. (2024). K-means centroids initialization based on differentiation between instances attributes. *International Journal of Intelligent Systems*, 2024(1), 7086878.
- Maahira, A. (2023). Analisis program pemberdayaan masyarakat e-katalog umkm untuk memajukan ekonomi masyarakat kota medan. *Jurnal Obor Penmas: Pendidikan Luar Sekolah*, 6(1), 51–60.
- Madhulatha, T. S. (2021). An overview on clustering methods. *IOSR Journal of Engineering*, 2(4), 719–725.
- Mawarni, H., Testiana, G., dan Dalafranka, M. L. (2023). Implementasi algoritma k-means untuk segmentasi pelanggan pada pt. bintang multi sarana cabang tugumulyo. *Jurnal Komputer dan Informatika*, 11(2), 227–236.
- Muditomo, A., dan Wahyudi, I. (2020). Conceptual model for sme digital transformation during the covid-19 pandemic time in indonesia: R-digital transformation model. , 3(1).

- Munawaroh, S. (2023). A review of data mining techniques in the development of decision support systems. *International Journal of Research and Applied Technology*, 2(3), 406–414.
- Oti, E. U., Olusola, M. O., Eze, F. C., dan Enogwe, S. U. (2021). Comprehensive review of k-means clustering algorithms. *International Journal of Advances in Scientific Research and Engineering*, 7(8), 64–69.
- Pawe, D. O. G., Utami, I., dan Man, S. (2024). Strategies for optimizing the use of local electronic catalogs in the procurement process. *Jurnal Ekonomi dan Bisnis*, 27(1), 1–20.
- Pérez-Ortega, J., Almanza-Ortega, N. N., Vega-Villalobos, A., Pazos-Rangel, R., Zavala-Díaz, C., dan Martínez-Rebollar, A. (2019). The k-means algorithm evolution. Dalam *Introduction to data science and machine learning*. IntechOpen.
- Pham, L.-H. T., Desai-Naik, T., Hammond, L., dan Abdeljabbar, W. (2021). *Information systems for business*.
- Prikustiawan, A. (2023). Analisis kebijakan pengadaan barang atau jasa melalui e-katalog lokal untuk pengembangan umkm kota surabaya. , 1(2).
- Radicic, D., dan Petković, S. (2023). Impact of digitalization on technological innovations in small and medium-sized enterprises (smes). *Technological Forecasting and Social Change*, 191, 122474.
- Ramadhan, A., Achmad, F., Zulkarnain, I., dan Aritsugi, M. (2025). Evaluation of k-means, dbscan, and hierarchical clustering for strategic segmentation of tourism smes in rembang, indonesia. *Jurnal Teknik Informatika (JUTIF)*, 6(3), 1605–1630.
- Ramadhan, H., Abdan Kamaludin, M. R., Nasrullah, M. A., dan Rolliawati, D. (2023). Comparison of hierarchical, k-means and dbscan clustering methods for credit card customer segmentation analysis based on expenditure level. *Journal of Applied Informatics and Computing*, 7(2), 246–251.
- Risco-Ramos, R. H., Pérez-Aguilar, D. A., Casaverde-Pacherrez, L. A., Malpica Rodríguez, M., Pérez-Aguilar, J. M., dan Pérez-Aguilar, A. P. (2023). Using a business intelligence framework and data mining as computational tools in smes: Production forecasting of a hydroelectric power plant as a case study. Dalam *Proceedings of the 21st laccei international multi-conference for engineering, education and technology*.
- Romero, I., dan Mammadov, H. (2024). Digital transformation of small and medium-sized enterprises as an innovation process: A holistic study of its

- determinants. *Journal of the Knowledge Economy*, 16(2), 8496–8523.
- Sajana, T., Sheela Rani, C. M., dan Narayana, K. V. (2023). A survey on clustering techniques for big data mining. *Indian Journal of Science and Technology*, 9(3). doi: 10.17485/ijst/2016/v9i3/75971
- Selayanti, N., Putri, S. A., dan Fahrudin, T. M. (2025). Analisis segmentasi sentra wisata kuliner untuk optimalisasi omzet umkm di surabaya menggunakan metode agglomerative hierarchical clustering. *JoMMiT: Jurnal Multi Media dan IT*, 8(2), 113–122. doi: 10.46961/jommit.v8i2.1351
- Setyowati, K., Prakoso, S. G., Utomo, I. H., Parwiyanto, H., Suryawati, R., Nugroho, R. A., ... Santoso, S. (2023). E-catalogue utilization counseling for small and medium enterprises association in surakarta city. *Dinamisia: Jurnal Pengabdian Kepada Masyarakat*, 7(6), 1744–1752.
- Sidhu, L. S., dan Sharma, J. (2023). Role of management information system. , 4(4).
- Simbolon, I. N., dan Friskila, P. D. (2024). Analisis dan evaluasi algoritma dbSCAN pada tuberkulosis. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3S1). doi: 10.23960/jitet.v12i3S1.5206
- Sinaga, M. H., Martina, S., dan Purba, D. (2024). Pengaruh modal kerja, jam kerja, dan tingkat pendidikan terhadap pendapatan umkm di kabupaten simalungun. *Jurnal Ilmiah Accusi*, 6(1), 151–160. doi: 10.36985/n4s0jb32
- Suci, Y. R. (2022). Perkembangan umkm di indonesia. , 6(1).
- Suryani, E., Nugroho, K. S., dan Cadith, J. (2025). The effect of e-purchasing implementation through e-catalogue on procurement performance. *Jurnal Bisnis Mahasiswa*, 5(4), 1753–1764. doi: 10.60036/jbm.674
- Syarifah, E., Purnamasari, S., dan Purnomo, A. (2021). Efektivitas penyaluran dana banpres produktif usaha mikro (bpm) untuk modal kerja dalam meningkatkan kesejahteraan pelaku umkm (studi kasus pada pelaku umkm di desa melayu). *Universitas Islam Kalimantan MAB*.
- Tertiaavini, T. (2024). A hybrid approach using k-means clustering and saw method for smes priority. *Journal of Intelligent System and Computation*, 6(1), 46–53. doi: 10.52985/insyst.v6i1.392
- Understanding sme heterogeneity: Towards policy relevant typologies* (Tech. Rep.). (2021). OECD. doi: 10.1787/c7074049-en
- Vankayalapati, R., Ghutugade, K. B., Vannapuram, R., dan Prasanna, B. P. S. (2021). K-means algorithm for clustering of learners performance levels using machine learning techniques. *Revue d'Intelligence Artificielle*, 35(1), 99–

104. doi: 10.18280/ria.350112
- Wang, S., Sun, Y., dan Bao, Z. (2020). On the efficiency of k-means clustering. *Proceedings of the VLDB Endowment*, 14(2), 163–175. doi: 10.14778/3425879.
- Wang, Y. (2023). Research on the evaluation of the growth level of smes clustering based on digitalization. Dalam X. Yuan, Y. Kurniawan, dan Z. Ji (Eds.), *Proceedings of the 2023 4th international conference on education, knowledge and information management (icekim 2023)* (Vol. 13, hal. 283–291). Atlantis Press International BV. doi: 10.2991/978-94-6463-172-2_31
- Wardani, S. D. K., Ariyanto, A. S., Umroh, M., dan Rolliawati, D. (2023). Perbandingan hasil metode clustering k-means, dbscan, dan hierarchical untuk analisa segmentasi pasar. *JIKO (Jurnal Informatika dan Komputer)*, 7(2), 191. doi: 10.26798/jiko.v7i2.796
- Wicaksana, D. A., Adikara, P. P., dan Adinugroho, S. (2018). Clustering dokumen skripsi dengan menggunakan hierarchical agglomerative clustering. , 2(12).
- Wijaya, F. B., Budiaji, W., dan Wicaksono, A. S. (2025). Applied machine learning dbscan for identifying clusters of micro and small industries. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 380–386. doi: 10.31004/riggs.
- Wikarta, N. C. A., dan Defiyanti, S. (2025). Pengelompokan umkm berdasarkan kategori usaha dan sebaran wilayah di jawa barat dengan clustering k-means. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3). doi: 10.23960/jitet.
- Xia, S., Peng, D., Meng, D., Zhang, C., Wang, G., Chen, Z., dan Wei, W. (2021). Ball k-means. *arXiv preprint arXiv:2005.00784*. doi: 10.48550
- Zhou, Q., dan Sun, B. (2024). Adaptive k-means clustering under-sampling methods. *Data and Information Management*, 8(3), 100064. doi: 10.1016/j.dim.2023.

LAMPIRAN A

Surat Izin Penelitian dan pengambilan data

KEMENTERIAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF SAMBAS RIAU
FAKULTAS SAINS DAN TEKNOLOGI
كلية العلوم والتكنولوجيا

Nomor : B-7262/F.V/PP.009/10/2025 Pekanbaru, 22 Oktober 2025
Sifat : Bista
Hal : Mohon Izin Penelitian dan Pengambilan Data
Tugas Akhir/Skripsi

Kepada Yth.

1. Kepala Layanan Penggandaan Secara Elektronik (LPSE)
Gedung Badan Pelayanan Terpadu (BPT) Lt. 3
Kantor Walikota Pekanbaru
Jl. Jend. Sudirman No. 464, Pekanbaru
2. Bagian Pengadaan Barang dan Jasa Kota Pekanbaru
Jl. Abdurrahman Hamid, Gedung Lontik
Kecamatan Tanjungpinang Raya, Kota Pekanbaru - Riau

Assalamualaikum Wr. Wb.
Dengan hormat, Sehubungan telah dimulainya mata kuliah Tugas Akhir pada Fakultas Sains dan Teknologi UIN Suska Riau, kami bermaksud menginikan mahasiswa:

Nama : Rahma Aliya
NIM : 12250320351
Fakultas : Sains dan Teknologi
Program Studi/Smt : Sistem Informasi /7
No. HdiC-mail : 082360463012

untuk pengambilan data yang sangat dibutuhkan dalam Tugas akhir mahasiswa tersebut yang berjudul "SEGMENTASI UMKM KOTA PEKANBARU BERDASARAN PARTISIPASI DALAM E-KATALOG LOKAL MENGGUNAKAN KMEANS CLUSTERING". Kami mohon kiranya saudara saudara berkenan memberikan izin dan fasilitas demi kelancaran Tugas Akhir mahasiswa yang bersangkutan.

Demikian surat ini kami sampaikan, atas bantuan dan kerjasama Saudara kami uscokan terimakasih.

Wissenskam
Dekan

Dr. Yuslenita Muda, M.Sc
NIP. 19770503200710000

Dokumen ini telah ditandai lengkap secara elektronik.
Tanda tangan: 

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN B

Surat Jawaban Izin Penelitian dan Pengambilan Data



UIN SUSKA RIAU

LAMPIRAN C

Dokumentasi Pengambilan Data di LPSE



© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN D

Dokumentasi Pengambilan Data di Dinas Koperasi



© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



DAFTAR RIWAYAT HIDUP

© Ha

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengujikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Penulis bernama Rahma Aliya, lahir di Kampung Jawa, Kecamatan Merbau, Kabupaten Kepulauan Meranti pada tanggal 02 November 2004. Penulis merupakan anak pertama dari tiga bersaudara perempuan, putri dari Ranu dan Maria Ulfah, dengan dua orang adik bernama Miftahul Jannah dan Zakiyah Ilma.

Pendidikan formal penulis dimulai pada usia lima tahun di TK Pertiwi. Selanjutnya, penulis menempuh pendidikan dasar di SD Negeri 13 Sepotong, Kecamatan Siak Kecil, Kabupaten Bengkalis pada tahun 2011–2017. Pendidikan menengah pertama diselesaikan di MTs GUPPI Bandar Sungai, Kecamatan Sabak Auh, Kabupaten Siak pada tahun 2017–2020.

Pada jenjang pendidikan menengah kejuruan, penulis melanjutkan studi di SMK Negeri 1 Siak Kecil, Kabupaten Bengkalis, pada jurusan Teknik Komputer dan Jaringan (TKJ) selama periode 2020–2023. Selama menempuh pendidikan di SMK, penulis mengikuti kegiatan Praktik Kerja Industri (Prakerin) selama enam bulan di PT. Wanxp Riau sebagai bagian dari penguatan kompetensi di bidang teknologi informasi.

Setelah menyelesaikan pendidikan menengah kejuruan, penulis melanjutkan pendidikan tinggi di Universitas Islam Negeri Sultan Syarif Kasim (UIN Suska) Riau pada tahun 2022 sebagai mahasiswa Program Studi Sistem Informasi jenjang Strata Satu (S1), Fakultas Sains dan Teknologi.

Selama masa perkuliahan, penulis aktif mengikuti berbagai kegiatan pengabdian kepada masyarakat di beberapa daerah melalui program kampus serta kolaborasi bersama dosen. Penulis juga terlibat dalam berbagai organisasi internal dan eksternal kampus. Pada organisasi internal, penulis merupakan anggota aktif Predatech dan tergabung dalam Divisi Public Relations. Selain itu, penulis memiliki ketertarikan dan pengalaman dalam publikasi ilmiah di bidang Sistem Informasi, khususnya pada topik data mining, audit sistem informasi, dan rekayasa perangkat lunak (RPL).