

Analisis Manfaat Penerapan *Machine Learning* untuk Otomatisasi Data *Engineering* di UMKM

Benefit Analysis of Machine Learning Implementation for Data Engineering Automation in MSMEs

¹Alexander Mario Saputra*, ²Magdalena A. Ineke Pakereng

^{1,2}Jurusan Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana

^{1,2}Jl. Diponegoro 52-60, Salatiga 50711, Indonesia

*e-mail: 672022026@student.uksw.edu

(received: 6 June 2026, revised: 20 June 2026, accepted: 21 June 2026)

Abstrak

Sektor usaha mikro, kecil, dan menengah menghadapi tekanan transformasi digital yang kian meningkat, namun masih terhambat oleh pengelolaan data yang bersifat manual sehingga rentan terhadap kesalahan dan inefisiensi operasional. Penelitian ini bertujuan mengkaji manfaat penerapan *machine learning* dalam mengotomatiskan proses *data engineering* di lingkungan usaha kecil, sekaligus memetakan hambatan adopsi dan mengidentifikasi celah penelitian yang belum terjawab. Pendekatan *systematic literature review* dengan kerangka Population, Intervention, Comparison, Outcome, Context (PICOC) diterapkan terhadap 25 artikel ilmiah terpilih terbitan 2021–2026 yang disaring dari enam basis data ilmiah, yaitu Google Scholar, ResearchGate, DOAJ, Garuda, Scopus, dan IEEE Xplore. Hasil sintesis menunjukkan empat manfaat utama, yakni berkurangnya ketergantungan pada pengolahan data manual, meningkatnya ketepatan prediksi bisnis, penghematan biaya operasional hingga 42% melalui percepatan aliran data, serta pengelolaan persediaan dan pelanggan yang lebih responsif. Adopsi teknologi ini masih terhambat oleh rendahnya kapasitas digital sumber daya manusia, keterbatasan infrastruktur dan finansial, serta data historis yang tidak memadai. Analisis distribusi literatur mengungkap bahwa 80% studi masih berpusat pada pemanfaatan keluaran *machine learning* di tataran operasional, sementara perancangan fondasi pengolahan data otomatis di lapisan hulu dalam konteks keterbatasan sumber daya usaha kecil hampir sepenuhnya terabaikan. Kekosongan inilah yang menjadi fokus utama penelitian ini.

Kata kunci: otomatisasi data, rekayasa data, machine learning, usaha mikro kecil menengah, tinjauan literatur sistematis

Abstract

Micro, Small, and Medium-Sized Enterprises (MSMEs) are facing increasing pressure to undergo digital transformation. However, many MSMEs continue to rely on manual data management processes, making them vulnerable to human error and operational inefficiencies. This study aims to examine the benefits of implementing machine learning to automate data engineering processes in MSMEs, while identifying the key barriers to adoption and highlighting unresolved research gaps in the existing literature. A Systematic Literature Review (SLR) was conducted using the Population, Intervention, Comparison, Outcome, and Context (PICOC) framework. A total of 25 peer-reviewed studies published between 2021 and 2026 were selected from six academic databases: Google Scholar, ResearchGate, DOAJ, Garuda, Scopus, and IEEE Xplore. The synthesis identified four major benefits of machine learning implementation: reduced dependence on manual data processing, improved business prediction accuracy, operational cost savings of up to 42% through more efficient data pipelines, and more responsive inventory and customer management. Despite these advantages, adoption remains constrained by limited digital capabilities of human resources, inadequate infrastructure and financial resources, and insufficient historical data for model development. Furthermore, the literature analysis revealed that approximately 80% of existing studies focus primarily on the operational use of machine learning outputs, whereas the design of automated upstream data engineering foundations tailored to the resource constraints of MSMEs has received

<http://sistemasi.ftik.unisi.ac.id>

very limited attention. This research gap represents the primary focus and contribution of the present study.

Keywords: *data automation, data engineering, machine learning, micro small medium enterprises, systematic literature review*

1 Pendahuluan

Arus transformasi digital yang bergerak semakin cepat telah mengubah lanskap persaingan bisnis secara fundamental, memaksa usaha mikro, kecil, dan menengah untuk turut beradaptasi dengan memanfaatkan teknologi informasi agar tidak tertinggal di tengah dinamika pasar yang terus berubah. Dalam ekosistem bisnis yang semakin berbasis data, kemampuan mengelola dan mengolah informasi secara efektif menjadi penentu daya saing yang tidak dapat diabaikan. Namun kenyataan di lapangan menunjukkan kondisi yang jauh dari ideal, di mana sebagian besar pelaku usaha kecil masih bergulat dengan keterbatasan infrastruktur teknologi, rendahnya literasi digital tenaga kerja, serta minimnya sumber daya manusia yang memiliki kompetensi memadai di bidang teknologi informasi [1].

Salah satu proses krusial yang menentukan kualitas pengelolaan data adalah *data engineering*, yang mencakup serangkaian aktivitas mulai dari pengumpulan, transformasi, integrasi, hingga penyimpanan data agar menghasilkan informasi yang terstruktur dan siap dianalisis. Ironisnya, sebagian besar usaha kecil masih menjalankan proses ini secara manual, kondisi yang tidak hanya memperlambat alur pengolahan data tetapi juga membuka celah kesalahan yang berulang dan sulit dikendalikan [2]. Minimnya pemanfaatan teknologi otomatisasi memperparah situasi ini dengan memperpanjang waktu pemrosesan dan menurunkan kualitas hasil analisis yang dihasilkan [3]. Kondisi ini secara tegas menunjukkan bahwa usaha kecil membutuhkan solusi teknologi yang mampu mengambil alih proses pengolahan data secara cerdas dan efisien.

Machine learning hadir sebagai jawaban strategis atas persoalan tersebut. Dengan kemampuannya mempelajari pola tersembunyi dari rekam jejak data historis, teknologi ini memungkinkan sistem untuk membuat prediksi dan menjalankan berbagai tahapan pemrosesan data secara otomatis tanpa memerlukan campur tangan manusia secara intensif [4]. Sejumlah kajian telah membuktikan bahwa penerapannya di lingkungan usaha kecil mampu mengoptimalkan pengelolaan rantai pasok, meningkatkan akurasi perkiraan permintaan, serta melakukan segmentasi pelanggan secara otomatis berdasarkan pola transaksi yang terekam [5]. Lebih jauh, otomatisasi berbasis *machine learning* juga terbukti mempercepat siklus analitik dan memperkuat fondasi pengambilan keputusan yang berpijak pada data [6].

Meski potensinya tidak diragukan, tingkat penetrasi *machine learning* dalam mengotomatiskan proses *data engineering* di kalangan usaha kecil masih sangat terbatas. Kurangnya pemahaman terhadap nilai strategis teknologi ini, sempitnya akses terhadap infrastruktur yang memadai, serta keterbatasan sumber daya untuk implementasi menjadi faktor-faktor yang secara bersamaan memperlambat adopsinya [7]. Kesenjangan antara potensi yang ditawarkan teknologi dan kemampuan nyata usaha kecil untuk mengadopsinya inilah yang memerlukan kajian mendalam dan berbasis bukti.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan: (1) mengidentifikasi manfaat konkret penerapan *machine learning* dalam mengotomatiskan proses *data engineering* pada usaha kecil; (2) memetakan hambatan dan tantangan utama yang menghambat adopsinya; serta (3) memetakan topik-topik yang belum terjawab oleh literatur yang ada sebagai pijakan bagi penelitian berikutnya. Secara akademis, penelitian ini berkontribusi memperkaya literatur di bidang *machine learning* dan *data engineering* khususnya dalam konteks usaha kecil di Indonesia. Secara praktis, temuan penelitian ini diharapkan menjadi panduan berbasis bukti bagi pelaku usaha, pembuat kebijakan, serta pengembang teknologi dalam merancang solusi otomatisasi data yang terjangkau dan sesuai kapasitas usaha kecil, guna mendukung percepatan transformasi digital sektor tersebut secara berkelanjutan.

2 Tinjauan Literatur

Sejumlah penelitian awal yang mengeksplorasi potensi *machine learning* (ML) di lingkungan UMKM umumnya bertolak dari kebutuhan mendesak untuk menggantikan proses pengelolaan data yang selama ini dilakukan secara manual dengan pendekatan yang lebih cerdas dan berbasis data.

Harahap et al. [8] mengungkapkan bahwa kecerdasan buatan secara umum telah membantu UMKM bertahan dan berkembang di era digital, meskipun penerapannya masih terbatas pada fungsi-fungsi bisnis tertentu. Hidayat dan Andarini [9] memperkuat hal ini dengan menunjukkan bahwa transformasi digital di lingkungan UMKM terbukti meningkatkan efisiensi pengelolaan bisnis secara nyata. Dalam konteks yang lebih spesifik, Hindrayani et al. [10] menemukan bahwa *machine learning* pada dasarnya mampu menopang berbagai kebutuhan analitik dasar UMKM, mulai dari pengelompokan konsumen berdasarkan pola perilaku hingga penetapan harga yang didasarkan pada rekam jejak transaksi historis, meskipun bukti penerapan nyata di lapangan masih sangat jarang dijumpai. Jeffry et al. [11] mendapati bahwa setelah mengikuti program pendampingan berbasis *machine learning*, pelaku usaha kerajinan rumahan berhasil membangun sistem prediksi dasar secara mandiri dengan peningkatan pemahaman teknologi yang signifikan. Kedua temuan ini memberi indikasi bahwa *machine learning* pada dasarnya dapat dijangkau oleh pelaku usaha kecil, namun keberhasilannya sangat bergantung pada ketersediaan dukungan eksternal yang terencana dan konsisten.

Sejumlah penelitian berikutnya bergerak lebih jauh dengan langsung menguji algoritma *machine learning* pada permasalahan operasional spesifik yang dihadapi UMKM. Yosephine dan Emor [6] mengambil pendekatan berbeda dengan memadukan kemampuan prediktif *machine learning* ke dalam perangkat sensor fisik sehingga peringatan ketersediaan stok dapat diberikan secara otomatis tanpa campur tangan operator. Widjaja dan Avinash [7] menemukan bahwa pendekatan berbasis *ensemble* dan klasifikasi vektor secara konsisten menghasilkan estimasi kebutuhan persediaan yang lebih cermat dibanding kalkulasi statistik konvensional. Basri et al. [12] melangkah lebih jauh dengan membangun model untuk mengelompokkan dan memetakan potensi usaha kecil di suatu wilayah dan berhasil mencapai tingkat presisi yang sangat tinggi. Meski ketiga studi ini menunjukkan hasil yang menjanjikan, semuanya hanya menyentuh satu fungsi bisnis secara terisolasi dan tidak mempersoalkan bagaimana data dari berbagai sumber seharusnya dikumpulkan, distandarisasi, dan dialirkan secara otomatis sebelum dapat diproses oleh model sebuah tantangan mendasar dalam *data engineering* yang justru menentukan kualitas keluaran secara keseluruhan.

Beberapa penelitian mulai bergerak ke lapisan yang lebih dalam dari sekadar penerapan model. Siswanto et al. [3] dalam kajian sistematis mereka menyimpulkan bahwa hingga kini belum ada rancangan sistem terintegrasi berbasis kecerdasan buatan yang sejak tahap perancangan awal benar-benar mempertimbangkan keterbatasan kapasitas khas yang dimiliki UMKM. Fathoni dan Bhrata [4] membangun sistem pengelolaan produksi pada usaha manufaktur skala kecil yang berhasil mengotomatiskan seluruh rangkaian pengolahan data tanpa perlu dikendalikan secara manual. Wahyudi et al. [5] mendemonstrasikan bahwa kemampuan *machine learning* yang ditanamkan langsung pada perangkat keras berdaya rendah dapat memantau kondisi stok secara terus-menerus dan mandiri tanpa keterlibatan operator. Banerjee [13] secara khusus meneliti bagaimana kecerdasan buatan dapat mengambil alih pengelolaan alur kerja *data engineering* secara menyeluruh dan menyimpulkan bahwa pendekatan ini secara substansial mempercepat siklus pemrosesan sekaligus menekan tingkat gangguan yang biasanya timbul dari keterlibatan manusia. Sridhar dan Kanani [14] memperkuat argumen ini dengan menegaskan bahwa kualitas dan kecepatan aliran data merupakan prasyarat yang tidak dapat diabaikan apabila suatu organisasi ingin membangun sistem *machine learning* yang benar-benar andal dalam jangka panjang. Agrawal et al. [15] mengembangkan kerangka kerja untuk mendeteksi dan memperbaiki kesalahan data secara otomatis, namun mengakui bahwa kemampuan sistemnya melemah secara signifikan ketika berhadapan dengan data berskala kecil dan tidak seimbang kondisi yang justru paling umum ditemui di lingkungan usaha kecil.

Dari sisi hambatan, berbagai penelitian secara konsisten mengidentifikasi bahwa tantangan adopsi *machine learning* di UMKM bersifat berlapis dan saling memperparah satu sama lain. Johan et al. [16] dan Ausat et al. [17] menyoroti bahwa minimnya penguasaan teknologi di kalangan tenaga kerja UMKM bukan sekadar hambatan teknis, melainkan kendala struktural yang mengakar jauh dan memerlukan waktu panjang untuk diatasi. Akadji [18] serta Nopitasari dan Fahdillah [19] mencatat bahwa kebutuhan modal awal untuk membangun ekosistem *machine learning* kerap jauh melampaui batas kemampuan keuangan sebagian besar pelaku usaha kecil. Giantri et al. [20] memperkirakan bahwa lebih dari dua pertiga organisasi di Indonesia menghadapi kesulitan nyata dalam mendapatkan tenaga ahli kecerdasan buatan yang memadai. Vajjhala dan Thandekkattu [21] memperluas cakupan hambatan ini dengan menunjukkan bahwa potensi ketidakadilan dalam keluaran model dan

<http://sistemasi.ftik.unisi.ac.id>

kerentanan terhadap kebocoran data pelanggan turut menjadi pertimbangan serius yang sering luput dari perhatian.

Bila ditelaah secara menyeluruh, pola yang terungkap dari seluruh literatur menunjukkan bahwa perhatian akademik terhadap *machine learning* di UMKM masih terpusat pada lapisan akhir pemanfaatan teknologi yakni bagaimana keluaran analitik digunakan untuk menunjang keputusan operasional sementara lapisan awal berupa pengumpulan, pembersihan, transformasi, dan pengintegrasian data secara otomatis hampir sepenuhnya luput dari kajian, khususnya dalam konteks keterbatasan infrastruktur dan sumber daya yang menjadi ciri khas UMKM. Belum ditemukan satu pun penelitian yang secara khusus merancang dan menguji secara empiris sebuah arsitektur *data engineering* berbasis *machine learning* yang mampu beroperasi efisien dengan data yang minim, tidak seragam, dan infrastruktur yang terbatas. Kekosongan inilah yang menjadi landasan dan arah utama penelitian ini, yaitu mengkaji secara sistematis sejauh mana *machine learning* dapat memberikan manfaat nyata dalam mengotomatiskan proses *data engineering* di lingkungan UMKM, sekaligus memetakan hambatan adopsinya guna membuka ruang bagi pengembangan solusi yang lebih realistis, ringan, dan kontekstual bagi sektor usaha kecil.

3 Metode Penelitian

Penelitian ini bersumber dari data sekunder berupa artikel ilmiah yang dihimpun melalui enam basis data daring sebagai tempat penelusuran utama, yaitu Google Scholar, ResearchGate, DOAJ, Garuda, Scopus, dan IEEE Xplore, dengan rentang waktu publikasi 2021 hingga 2026. Pengelolaan dan deduplikasi referensi dilakukan menggunakan aplikasi Mendeley, sementara proses seleksi dicatat secara rinci melalui lembar kerja berbasis spreadsheet. Bahan utama penelitian mencakup protokol SLR yang disusun sebelum pencarian dimulai dan formulir ekstraksi data terstandar tujuh item sebagai instrumen analisis setiap artikel yang lolos seleksi.

Pendekatan yang diterapkan adalah *Systematic Literature Review* (SLR) menggunakan kerangka Population, Intervention, Comparison, Outcome, Context (PICOC) sebagai panduan perumusan pertanyaan penelitian, dengan rancangan kegiatan empat tahap berurutan: perumusan pertanyaan penelitian, pencarian dan pengumpulan artikel, penyaringan kelayakan, serta ekstraksi dan sintesis data [13]. Ruang lingkup mencakup studi yang membahas penerapan *machine learning* dalam proses *data engineering* meliputi ETL, *data pipeline*, pembersihan data, dan deteksi anomali pada usaha mikro, kecil, dan menengah di Indonesia maupun negara berkembang lainnya. Penjabaran komponen PICOC disajikan pada Tabel 1.

Tabel 1 Komponen kerangka PICOC penelitian

Komponen	Deskripsi
Population	Usaha Mikro, Kecil, dan Menengah (UMKM)
Intervention	Penerapan ML dalam proses <i>Data Engineering</i> (ETL, <i>data pipeline</i> , <i>data cleaning</i> , <i>anomaly detection</i>)
Comparison	Proses <i>Data Engineering</i> manual tanpa otomatisasi berbasis ML
Outcome	Manfaat operasional, tantangan adopsi, dan celah penelitian yang teridentifikasi
Context	Lingkungan UMKM dengan keterbatasan infrastruktur dan SDM

Berdasarkan kerangka PICOC tersebut, tiga *Research Question* (RQ) dirumuskan sebagai berikut:

RQ1: Apa saja manfaat yang diperoleh UMKM dari penerapan *Machine Learning* dalam otomatisasi proses *Data Engineering*?

RQ2: Apa saja hambatan dan tantangan utama yang dihadapi UMKM dalam mengadopsi *Machine Learning* untuk otomatisasi *Data Engineering*?

RQ3: Apa celah penelitian yang masih terbuka berdasarkan literatur yang telah dianalisis?

Pengumpulan data dilakukan melalui penelusuran sistematis dengan kata kunci: ("Machine Learning" OR "ML") AND ("Data Engineering" OR "Data Pipeline" OR "ETL") AND ("UMKM" OR "SME"). Penyaringan dilakukan dua tahap berdasarkan kriteria pada Tabel 2, yaitu seleksi judul dan abstrak, dilanjutkan pembacaan *full-text* dan penilaian kualitas menggunakan *quality checklist*.

Tabel 2 Kriteria inklusi dan eksklusi

Aspek	Kriteria Inklusi	Kriteria Eksklusi
Periode	Tahun 2021–2026	Sebelum tahun 2021
Bahasa	Indonesia atau Inggris	Selain kedua bahasa tersebut
Topik	ML dan/atau otomatisasi data dalam konteks UMKM	Tidak berkaitan dengan UMKM atau ML
Jenis Publikasi	Jurnal atau prosiding Terindeks	Opini, editorial, buku teks, tesis
Kualitas	<i>Full-text</i> tersedia secara terbuka	Hanya abstrak atau artikel berbayar
	Metodologi jelas dan temuan terverifikasi	Metodologi tidak jelas atau tidak terverifikasi

Dari 47 artikel hasil identifikasi awal, setelah deduplikasi tersisa 41, disaring menjadi 34 berdasarkan judul dan abstrak, kemudian 28 setelah pembacaan *full-text*, dan akhirnya diperoleh 25 artikel final setelah penilaian kualitas sebagaimana ditunjukkan pada Tabel 3.

Tabel 3 Alur Seleksi Artikel

Tahap	Deskripsi	Jumlah
Identifikasi awal	Penelusuran dari 6 basis data	47
Deduplikasi	Penghapusan artikel duplikat	41
Penyaringan judul & abstrak	Seleksi relevansi topik dan periode	34
Pembacaan full-text	Verifikasi kesesuaian isi	28
Penilaian kualitas	Seleksi berdasarkan <i>quality checklist</i>	25

Variabel *manfaat* didefinisikan sebagai dampak positif terukur dari penerapan ML pada *Data Engineering* UMKM, mencakup efisiensi, akurasi, kecepatan, dan penghematan biaya. *Hambatan* merujuk pada faktor teknis, finansial, maupun SDM yang terbukti menghambat adopsi ML. Sedangkan *celah penelitian* adalah topik yang belum dikaji atau kesenjangan antara temuan literatur dengan kebutuhan lapangan. Analisis dilakukan dengan mengelompokkan temuan dari seluruh artikel berdasarkan tema yang relevan, kemudian dibandingkan untuk menemukan pola dan perbedaan antar studi.

4 Hasil dan Pembahasan

Bagian ini menguraikan hasil sintesis terhadap 25 artikel terpilih secara berurutan sesuai tahapan metode yang telah ditetapkan, mulai dari hasil perumusan pertanyaan penelitian, proses seleksi literatur, karakteristik demografis artikel yang terpilih, hingga temuan substantif yang menjawab ketiga RQ dan posisi penelitian ini di antara studi-studi sebelumnya.

Hampir separuh artikel terbit pada periode 2025–2026 (48%), diikuti 2023–2024 (40%), dan 2021–2022 (12%), mencerminkan akselerasi minat akademik terhadap *machine learning* di sektor usaha kecil dalam dua tahun terakhir. Dari sisi fokus, 52% studi membahas *machine learning* pada fungsi operasional bisnis, 28% mengkaji transformasi digital UMKM secara makro, dan hanya 20% yang secara spesifik menyoroti otomatisasi *data engineering* di lapisan hulu [3] [13] [14]. Karakteristik selengkapnya disajikan pada Tabel 4.

Tabel 4 Karakteristik demografis 25 artikel yang dianalisis

Karakteristik	Kategori	Jumlah (%)
Tahun Publikasi	2021–2022	3 (12%)
	2023–2024	10 (40%)
	2025–2026	12 (48%)
Fokus Utama	Operasional bisnis	13 (52%)
	Transformasi digital UMKM	7 (28%)
	<i>Data Engineering</i> / level hulu	5 (20%)
Algoritma ML	Random Forest, SVM	8 (32%)
	AI/ML umum	9 (36%)
	XGBoost, AutoML, IoT+ML, <i>Deep Learning</i>	8 (32%)
	Nasional (SINTA)	
Asal Jurnal	Internasional & Prosiding	18 (72%)
		7 (28%)

RQ1: Manfaat Penerapan ML dalam Otomatisasi Data Engineering pada UMKM

Penelusuran terhadap seluruh literatur menghasilkan empat kelompok dampak positif yang secara berulang dikonfirmasi dari berbagai konteks penerapan yang berbeda. Kelompok pertama berkaitan dengan pengurangan ketergantungan pada pekerjaan manual dalam pengolahan data. Fathoni dan Bhrata [4] merancang sebuah sistem cerdas pada usaha produksi skala kecil yang berhasil mengambil alih seluruh rangkaian pemrosesan data dari tangan manusia ke mekanisme otomatis berbasis analitik. Hasil serupa ditunjukkan oleh Wahyudi et al. [5] yang menanamkan kecerdasan komputasional langsung pada perangkat keras bertenaga rendah sehingga kondisi stok dapat dipantau secara mandiri tanpa jeda. Banerjee [13] pada tataran yang lebih luas menyimpulkan bahwa sistem berbasis kecerdasan buatan yang mengelola alur kerja pengolahan data secara otonom terbukti mempercepat siklus pemrosesan sekaligus menekan frekuensi gangguan yang lazim ditimbulkan oleh keterlibatan manusia.

Kelompok kedua mencakup peningkatan ketepatan perkiraan dan mutu informasi yang digunakan sebagai dasar keputusan bisnis. Widjaja dan Avinash [7] membandingkan beberapa pendekatan algoritmik dan menemukan bahwa metode berbasis pembelajaran mesin secara konsisten menghasilkan prakiraan kebutuhan barang yang lebih presisi dibanding teknik perhitungan tradisional. Basri et al. [12] bahkan berhasil melampaui angka presisi 98% ketika membangun model pengelompokan potensi usaha kecil di suatu daerah. Rahardi et al. [22] turut mempertegas hal ini dengan membuktikan bahwa pendekatan berbasis gabungan pohon keputusan paling unggul dalam memetakan tingkat kepuasan konsumen pada platform transaksi daring milik usaha kecil.

Kelompok ketiga menyangkut percepatan pergerakan data yang secara langsung berdampak pada efisiensi pengeluaran operasional. Giantri et al. [20] melaporkan bahwa penyatuan sistem kecerdasan buatan dengan infrastruktur komputasi berbasis awan mampu mendongkrak ketepatan proyeksi permintaan hingga 37% sekaligus memangkas beban biaya operasional dalam kisaran 28% hingga 42%. Sridhar dan Kanani [14] menambahkan bahwa jalur aliran data yang dipercepat dan dioptimalkan melalui otomatisasi turut mengeliminasi pemborosan sumber daya yang selama ini tercipta akibat proses pengolahan yang berlapis dan memakan waktu lama.

Kelompok keempat berkenaan dengan kemampuan adaptasi sistem pengelolaan persediaan dan hubungan pelanggan terhadap perubahan kondisi pasar yang dinamis. Yosephine dan Emor [6] menunjukkan bahwa integrasi sensor fisik dengan kecerdasan komputasional menghasilkan mekanisme peringatan ketersediaan barang yang berjalan sepenuhnya otomatis, memungkinkan pelaku usaha merespons fluktuasi permintaan jauh lebih cepat dari sebelumnya. Diantika dan Firmanto [23] melengkapi temuan ini dengan membuktikan bahwa sistem saran produk yang digerakkan oleh algoritma secara otomatis terbukti mendorong peningkatan angka konversi pada kanal penjualan daring milik usaha kecil. Keseluruhan kelompok manfaat tersebut beserta indikator dan studi pendukungnya dirangkum pada Tabel 5.

Tabel 5 Pemetaan manfaat utama penerapan ML dalam

data engineering pada UMKM (RQ1)

Manfaat Utama	Indikator Terukur	Studi Pendukung
Efisiensi operasional	Reduksi proses manual; penurunan latensi pemrosesan	[4], [5], [13]
Akurasi prediksi	Presisi model hingga 98,63%	[7], [12], [22]
Penghematan biaya	Penurunan biaya 28–42%; peramalan naik 37%	[14], [15], [20]
Responsivitas pasar	Notifikasi stok otomatis; rekomendasi pelanggan <i>real-time</i>	[6], [23], [24]

RQ2: Hambatan dan Tantangan Adopsi ML pada UMKM

Tiga gugus hambatan berulang kali muncul secara konsisten lintas studi. Pertama, minimnya kapasitas digital sumber daya manusia sebagai rintangan paling mendasar. Johan et al. [16] dan Ausat et al. [17] menegaskan bahwa rendahnya penguasaan teknologi tenaga kerja UMKM merupakan hambatan sistemik yang memerlukan waktu panjang untuk diatasi, diperkuat oleh temuan Giantri et al. [20] bahwa sekitar dua pertiga organisasi di Indonesia kesulitan mendapatkan tenaga ahli kecerdasan buatan.

Kedua, sempitnya ruang gerak akibat keterbatasan infrastruktur dan keuangan. Nopitasari dan Fahdillah [19] mencatat kebutuhan modal awal *machine learning* kerap jauh melampaui kemampuan finansial UMKM, sementara Banerjee [13] menambahkan bahwa sistem teknologi lama yang lazim dipakai UMKM umumnya tidak kompatibel dengan solusi *machine learning* terkini.

Ketiga, ketidakcukupan data historis dari segi jumlah maupun mutu. Widjaja dan Avinash [7] mencatat rekam jejak transaksi UMKM terlalu terbatas untuk melatih model yang andal. Agrawal et al. [15] membuktikan akurasi sistem deteksi anomali merosot tajam pada data kecil dan tidak proporsional kondisi umum di lingkungan UMKM sedangkan Vajjhala dan Thanekkattu [21] menyoroti risiko bias model dan kebocoran data pelanggan sebagai tantangan tambahan yang sering diabaikan. Ketiga gugus ini dirangkum pada Tabel 6.

Tabel 6 Pemetaan hambatan adopsi ML pada UMKM (RQ2)

Klaster Hambatan	Studi Pendukung	Dampak
SDM & Literasi Digital	[9], [16], [17], [20]	Sulit operasikan sistem ML; 67% organisasi kesulitan rekrut talenta AI
Infrastruktur & Finansial	[4], [6], [13], [18], [19]	Biaya tinggi; sistem lama tidak kompatibel solusi ML modern
Kualitas & Kuantitas Data	[7], [15], [21], [25]	Data terbatas menurunkan akurasi; risiko bias dan privasi data

RQ3: Celah Penelitian yang Teridentifikasi

Pola yang terungkap dari keseluruhan proses analisis memperlihatkan ketimpangan yang cukup mencolok dalam sebaran perhatian penelitian. Dari total 25 artikel yang dikaji, hanya lima studi yang secara eksplisit menempatkan otomatisasi *data engineering* di lapisan hulu sebagai objek kajian utamanya, yakni Siswanto et al. [3], Fathoni dan Bhrata [4], Banerjee [13], Sridhar dan Kanani [14], serta Agrawal et al. [15]. Selebihnya, yakni delapan puluh persen dari keseluruhan literatur, masih berfokus pada pemanfaatan keluaran analitik *machine learning* untuk keperluan operasional sehari-hari tanpa menyentuh persoalan bagaimana fondasi pengolahan data yang menopang seluruh sistem tersebut seharusnya dibangun, dikelola, dan dirawat. Distribusi ketimpangan ini digambarkan secara rinci pada Tabel 7.

Tabel 7 Distribusi fokus penelitian terhadap level penerapan ML (RQ3)

Level Penerapan	Jumlah	Persentase	Contoh Studi
-----------------	--------	------------	--------------

Hilir, operasional bisnis	13	52%	[6], [7], [12], [22],
Makro, transformasi digital	7	28%	[23]
Hulu, <i>Data Engineering & pipeline</i>	5	20%	[8], [9], [16], [17], [20] [3], [4], [13], [14], [15]

Dari ketimpangan tersebut, tiga kekosongan penelitian yang paling mendesak untuk segera diisi dapat dipetakan. Pertama, hingga saat ini belum ada satu pun rancangan arsitektur *data engineering* berbasis *machine learning* yang sejak awal didesain secara khusus untuk beroperasi dalam kondisi keterbatasan infrastruktur dan sumber daya yang menjadi ciri khas lingkungan UMKM [3]. Kedua, tidak dijumpai satu pun kajian yang secara langsung membandingkan berbagai pendekatan otomatisasi *pipeline* data dalam situasi nyata usaha kecil di Indonesia, sehingga rekomendasi yang ada masih belum berpijak pada bukti empiris yang kuat [14]. Ketiga, sebagian besar solusi yang telah dikembangkan oleh peneliti hingga kini masih diujicobakan pada kumpulan data simulasi atau pada organisasi dengan skala dan sumber daya yang jauh lebih besar, sehingga tingkat kesesuaiannya dengan kondisi data UMKM yang minim, tidak terstruktur, dan tidak seragam belum dapat dipastikan secara meyakinkan [15] [21].

Dibanding kajian-kajian sebelumnya, penelitian ini menonjol dalam empat hal. Pertama, menempatkan *data engineering* sebagai objek utama tinjauan dalam konteks UMKM bukan sekadar aplikasi *machine learning* di level bisnis menjadikannya salah satu yang pertama mengisi celah yang selama ini kurang diperhatikan. Kedua, kerangka PICOC menghasilkan batasan kajian yang terukur dan dapat direproduksi, melampaui tinjauan naratif yang bergantung pada penilaian subjektif. Ketiga, penelusuran lintas enam basis data menghasilkan cakupan literatur yang lebih menyeluruh dibanding studi serupa. Keempat, sintesis terpadu antara manfaat, hambatan, dan celah penelitian menghasilkan peta yang dapat langsung ditindaklanjuti oleh peneliti maupun praktisi kontribusi yang belum dijumpai pada studi-studi dengan tema serupa sebelumnya.

5 Kesimpulan

Kajian sistematis terhadap 25 artikel ilmiah terbitan 2021–2026 mengungkap bahwa pemanfaatan *machine learning* untuk mengotomatiskan proses *data engineering* khususnya ekstraksi dan transformasi data, pembersihan data otomatis, pengelolaan *data pipeline*, serta deteksi anomali memberikan dampak nyata bagi UMKM dalam empat wujud: berkurangnya ketergantungan pada pengolahan data manual [4] [5] [13], meningkatnya ketepatan prediksi bisnis dengan presisi terbaik mencapai 98,63% [7] [12] [22], terpengkasnya biaya operasional hingga 42% melalui percepatan aliran data [14] [15] [20], serta sistem pengelolaan persediaan dan pelanggan yang lebih tanggap terhadap perubahan pasar [6] [23] [24]. Kendati demikian, adopsi teknologi ini masih terbentur tiga rintangan utama yang saling memperparah, yaitu rendahnya kapasitas digital tenaga kerja sebagai hambatan paling mendasar [16] [17] [20], ketidaksesuaian antara besarnya investasi infrastruktur dengan kemampuan finansial UMKM [13] [18] [19], serta rekam jejak data historis yang terlalu terbatas dan tidak seragam untuk melatih model yang andal [7] [15] [21]. Analisis distribusi fokus penelitian mengungkap bahwa 80% literatur yang ada masih berkutat pada pemanfaatan keluaran *machine learning* di tataran operasional bisnis, sementara perancangan fondasi pengolahan data secara otomatis di lapisan hulu hampir sepenuhnya diabaikan dalam konteks keterbatasan sumber daya UMKM [3] [13] [14], sehingga membuka peluang bagi penelitian selanjutnya untuk merancang dan menguji secara empiris arsitektur *data engineering* berbasis *machine learning* yang ringan, hemat sumber daya, dan benar-benar disesuaikan dengan kondisi nyata UMKM sebagai pijakan transformasi digital sektor usaha kecil yang lebih kokoh dan berkelanjutan.

Referensi

- [1] N. Ruseno, S. Kurniawan, and G. Santoso, "Pemanfaatan *Big Data*, *AI*, dan *Robotik* untuk Prediksi Tren Ekonomi berbasis *Machine Learning*," *Jurnal PASTI*, Vol. 5, No. 1, pp. 18–29, 2025.

- [2] Y. Yeni and M. K. Darmaputera, "Mengeksplorasi Kecerdasan Buatan pada Manajemen Pemasaran Digital Era 5.0 di Dunia UMKM," *Transekonomika: Akuntansi, Bisnis dan Manajemen*, Vol. 5, No. 2, pp. 45–57, 2024.
- [3] T. Siswanto, S. Sari, and H. Hartini, "Integrasi Artificial Intelligence pada Aplikasi ERP: Systematic Literature Review," *Jurnal Teknik Informatika UST*, Vol. 7, No. 1, pp. 1–9, 2024.
- [4] M. A. Fathoni and A. A. Bhrata, "Smart Manufacturing Management System memanfaatkan Big Data dan Algoritma Machine Learning untuk Produksi UMKM," *SINTECH*, Vol. 5, No. 2, pp. 104–111, 2022.
- [5] W. Wahyudi, A. Hidayat, and S. W. Putra, "Penerapan Machine Learning pada Mikrokontroler Arduino Mega PRO MINI ATmega2560-16AU," *Jurnal Embedded System & Security*, Vol. 2, No. 1, pp. 31–38, 2022.
- [6] V. S. Yosephine and G. N. K. Emor, "Perancangan Sistem Pendeteksi Stok berbasis Machine Learning dan Mikrokontroler untuk Digitalisasi UMKM," *Jurnal Serambi Engineering*, Vol. 4, No. 1, pp. 89–97, 2024.
- [7] A. Widjaja and A. Avinash, "Analisis Perbandingan Algoritma Machine Learning untuk Forecasting Persediaan Produk Barang Pokok," *Jurnal Teknik Informatika - JUTISI*, Vol. 10, No. 2, pp. 203–212, 2024.
- [8] A. R. Harahap, C. Munthe, M. F. Hariri, and S. P. Lubis, "Peran AI dalam UMKM: Bagaimana kecerdasan buatan membantu UMKM Bertahan dan Berkembang di Era Digital," *IKRAITH-EKONOMIKA*, Vol. 8, No. 1, pp. 409–419, 2025.
- [9] T. Hidayat and R. M. Andarini, "Pelaksanaan Transformasi Digital dalam Pengelolaan Bisnis UMKM di Wilayah Kerek," *Jurnal Media Akademik*, Vol. 3, No. 2, pp. 1–12, 2025.
- [10] K. M. Hindrayani, A. Anjani, and A. L. Nurlaili, "Penerapan Machine Learning pada Penjualan Produk UMKM: Studi Literatur," *Prosiding Seminar Nasional Sains Data*, Vol. 1, No. 1, pp. 19–23, 2021.
- [11] Jeffrey, Nurdyansa, S. Usman, D. Sasmita, and M. N. Arafah, "Transformasi Digital dalam Industri Kerajinan Rumahan melalui Penggunaan Teknologi Machine Learning," *JMM (Jurnal Masyarakat Mandiri)*, Vol. 7, No. 6, pp. 5840–5849, 2023.
- [12] Basri, R. Rachman, Z. Said, and R. Idrus, "Model Integrasi Machine Learning dan Decision Support System dalam Pemetaan Potensi UMKM Kabupaten Polewali Mandar," *Bulletin of Information Technology*, Vol. 7, No. 1, pp. 50–58, 2026.
- [13] A. Banerjee, "Automating Data Engineering Workflows with AI and Machine Learning," *International Journal of Artificial Intelligence, Data Science and Machine Learning*, Vol. 5, No. 2, pp. 9–16, 2024.
- [14] R. Sridhar and I. J. Kanani, "The Role of Data Engineering in Enabling Machine Learning and Artificial Intelligence," *International Journal of Artificial Intelligence, Data Science and Machine Learning*, Vol. 5, No. 4, pp. 214–226, 2024.
- [15] J. Agrawal, V. K. T. Thakur, and S. Thakur, "A Machine Learning Framework for Automated Data Cleaning and Anomaly Detection in Large Datasets," *IJSRCSE*, Vol. 13, No. 3, pp. 21–29, 2025.
- [16] H. P. R. P. A. Johan and V. L. Wati, "Transformasi Digital UMKM: Menggali Peluang dan Tantangan Era Online Shop di Indonesia," *Triwikrama: Jurnal Ilmu Sosial*, Vol. 7, No. 3, pp. 51–59, 2025.
- [17] A. M. A. Ausat, D. O. Suparwata, and A. Risdwiyanto, "Optimalisasi Digital Competence sebagai Strategi Adaptasi Dinamis Wirausahawan dalam menghadapi Disrupsi Pasar di Era Digital," *Jurnal Minfo Polgan*, Vol. 14, No. 1, pp. 1–10, 2025.
- [18] I. Akadji, "Integrasi Sistem Informasi Manajemen dengan Artificial Intelligence untuk Prediksi Keputusan Bisnis pada UMKM," *Journal of Innovative and Creativity*, Vol. 5, No. 2, pp. 7976–7985, 2025.
- [19] M. R. Nopitasari and Y. Fahdillah, "Implementasi Transformasi Digital dan Kecerdasan Buatan sebagai Inovasi untuk UMKM pada Era Revolusi Industri 4.0," *Jurnal Ilmiah Manajemen dan Kewirausahaan (JUMANAGE)*, Vol. 3, No. 1, pp. 266–273, 2024.
- [20] L. T. Giantri et al., "Optimizing Digital Transformation through AI and Cloud Technology Integration for Innovation in Big Data-Driven Information Systems," *JAIC*, Vol. 9, No. 5, pp. 2652–2662, 2025.

- [21] N. R. Vajjhala and S. G. Thandekkattu, "*Machine Learning in Small- and Medium-Sized Enterprises' Knowledge Management: An Overview of the Latest Trends*," *Springer Proceedings in Mathematics & Statistics*, Vol. 492, pp. 301–315, 2025.
- [22] A. Rahardi et al., "Pemanfaatan *Machine Learning* untuk Prediksi Kepuasan Pelanggan pada UMKM Digital," *Journal of Data Science Methods and Applications*, Vol. 1, No. 2, pp. 66–74, 2025.
- [23] A. S. Diantika and Y. Firmanto, "Implementasi *Machine Learning* pada Aplikasi Penjualan Produk Digital (Studi pada GrabKios)," *Jurnal Ilmiah Mahasiswa FEB Universitas Brawijaya*, Vol. 13, No. 1, pp. 1–15, 2025.
- [24] Pratama et al., "Pemanfaatan *Business Intelligence* dalam UMKM: Tinjauan Literatur terhadap Peluang dan Tantangan Transformasi Data-Driven," *Indonesian Journal of Digital Business*, Vol. 2, No. 1, pp. 1–18, 2025.
- [25] Y. A. Saputra and D. Mahdiana, "Analisis Penerapan *Machine Learning*, *Deep Learning*, dan *Data Mining* dalam Prediksi Penjualan di Industri Otomotif," *Jurnal Pendidikan dan Teknologi Indonesia*, Vol. 5, No. 6, pp. 1743–1755, 2025.