JOURNAL OF HUMAN AND EDUCATION
E-ISSN: 2776-5857 | P-ISSN: 2776-7876
Website: https://jahe.or.id/index.php/jahe
Vol. 6, No. 6 (2026)

Transformasi Peramalan Permintaan (*Demand Forecasting*) Dan Pengendalian Persediaan Berbasis *Artificial Intelligence*: Kerangka Konseptual Efisiensi Biaya Operasional

Suhardi, Nur Indah, Suriyanti, Ramlawati
Pascasarjana Universitas Muslim Indonesia
**Email: suhardi.18071997@gmail.com*

ABSTRAK
Ketidakpastian permintaan, efek cambuk, dan trade-off antara biaya simpan dan biaya kekurangan stok membuat peramalan statistik tradisional dan model persediaan statis kurang memadai. Studi konseptual ini merumuskan kerangka efisiensi biaya operasional melalui sintesis literatur sistematis atas 28 artikel (2019–2026) yang dihimpun dengan Publish or Perish dan dianalisis secara deduktif-komparatif. Temuan konseptual berupa kerangka AI-Driven Demand & Inventory Optimization (AIDIO) yang mengaitkan penurunan galat peramalan dengan penyesuaian dinamis parameter persediaan. Implikasinya, ritel dan UMKM Indonesia perlu mengadopsi AI secara bertahap sesuai kesiapan data.
Kata Kunci: biaya operasional, efek cambuk, kecerdasan buatan, pengendalian persediaan, peramalan permintaan

ABSTRACT
Demand uncertainty, the bullwhip effect, and the trade-off between holding and shortage costs make traditional statistical forecasting and static inventory models inadequate. This conceptual study develops an operational cost-efficiency framework through a systematic literature synthesis of 28 articles (2019–2026), retrieved with Publish or Perish and analysed deductively and comparatively. The AI-Driven Demand & Inventory Optimization (AIDIO) framework links lower forecast error to dynamic inventory-parameter adjustment. Indonesian retailers and MSMEs should adopt AI gradually, according to data readiness.
Keywords: artificial intelligence, bullwhip effect, demand forecasting, inventory control, operational cost

PENDAHULUAN

Manajemen operasional modern berada di bawah tekanan untuk memenuhi permintaan yang makin sulit diprediksi dengan persediaan yang makin ramping. Persediaan berfungsi sebagai penyangga antara ketidakpastian permintaan dan ketidakpastian pasokan, tetapi penyangga itu tidak gratis. Persediaan berlebih (*overstock*) menimbulkan biaya simpan (*holding cost*) berupa biaya modal yang tertahan, ruang gudang, asuransi, penyusutan, dan risiko keusangan atau kerusakan pada barang yang mudah rusak. Sebaliknya, persediaan yang terlalu tipis menimbulkan biaya kekurangan stok (*stockout*/*shortage cost*) berupa penjualan yang hilang, pesanan tertunda, dan menurunnya loyalitas pelanggan. Kedua biaya ini bergerak berlawanan arah sehingga setiap keputusan persediaan pada dasarnya adalah keputusan menyeimbangkan keduanya di bawah ketidakpastian (Ünal dkk., 2023; Li dkk., 2025).

Ketidakpastian itu tidak berhenti pada satu perusahaan. Lee dkk. (1997) menunjukkan bahwa varians pesanan membesar ketika bergerak ke hulu rantai pasok, fenomena yang dikenal

sebagai efek cambuk (*bullwhip effect*). Penyebab utamanya adalah pembaruan peramalan permintaan yang dilakukan setiap pihak secara terpisah, pemesanan dalam lot besar, fluktuasi harga, dan penjatahan pasokan. Dengan demikian, mutu peramalan di tingkat hilir menentukan seberapa besar distorsi informasi yang diteruskan ke pemasok. Peramalan yang buruk membuat setiap mata rantai menimbun *safety stock* lebih besar daripada yang sebenarnya diperlukan, dan biaya sistemik yang timbul jauh melampaui biaya satu gudang.

Praktik peramalan dan pengendalian persediaan yang masih dominan bertumpu pada metode statistik klasik. *Moving average*, *exponential smoothing*, dan ARIMA konvensional (Box & Jenkins, 1970) mengasumsikan struktur linier dan stasioner, sehingga kurang tangguh menghadapi promosi, pola musiman yang berubah, permintaan sporadis, dan guncangan eksternal. Studi komparatif pada perusahaan alat berat menunjukkan perlunya membandingkan regresi dengan metode kecerdasan buatan (*artificial intelligence*/AI) pada data permintaan nyata (Aktepe dkk., 2021), sementara tinjauan kritis memperlihatkan bahwa model *machine learning* dan *deep learning* kini menjadi arus utama peramalan permintaan rantai pasok (Douaioui dkk., 2024; Terrada dkk., 2022). Pada sisi persediaan, model kuantitas pesanan ekonomis atau EOQ (Harris, 1913) mengandaikan laju permintaan tetap dan parameter biaya konstan, sedangkan *safety stock* konvensional umumnya memakai deviasi standar historis yang statis. Dalam lingkungan non-linear dan dinamis, kedua asumsi tersebut menghasilkan parameter yang cepat usang.

Meskipun literatur AI untuk persediaan berkembang pesat, terdapat celah penelitian yang penting. Pertama, tinjauan yang ada cenderung terpisah: sebagian berfokus pada metode peramalan (Atwani dkk., 2024; Douaioui dkk., 2024), sebagian pada aplikasi AI dalam pengelolaan persediaan (Ünal dkk., 2023; Hierro dkk., 2026). Kedua, studi yang mengintegrasikan peramalan dan kebijakan persediaan (Seyedan dkk., 2023; Yang dkk., 2025) umumnya bersifat teknis dan belum diterjemahkan menjadi kerangka manajerial yang menghubungkan akurasi prediktif dengan biaya operasional. Ketiga, dalam korpus yang dianalisis, kajian berkonteks Indonesia masih terbatas pada rantai pasok UMKM tekstil (Khawarizmi dkk., 2026) dan studi kualitatif pelaku usaha (Purba dkk., 2026), padahal ritel dan UMKM merupakan segmen yang paling terpapar risiko kelebihan dan kekurangan stok.

Berdasarkan celah tersebut, artikel ini bertujuan: (1) memetakan evolusi metode peramalan permintaan dari model statistik linier menuju analitik prediktif berbasis AI; (2) membandingkan paradigma operasi konvensional dan paradigma operasi berbasis AI pada aspek peramalan, penentuan parameter persediaan, mitigasi efek cambuk, dan biaya; (3) menjelaskan mekanisme reduksi biaya persediaan melalui akurasi prediktif; dan (4) merumuskan kerangka konseptual *AI-Driven Demand & Inventory Optimization* (AIDIO) beserta implikasi manajerialnya bagi korporasi ritel dan UMKM di Indonesia.

KAJIAN TEORETIS

Evolusi Metode Peramalan Permintaan menuju *Predictive Analytics* Berbasis AI

Evolusi peramalan permintaan dapat dibaca sebagai pergeseran dari model yang menjelaskan masa depan dengan satu deret waktu menuju model yang mempelajari pola dari banyak sumber data. Generasi pertama adalah model statistik linier univariat, yaitu *moving average*, *exponential smoothing*, dan ARIMA (Box & Jenkins, 1970). Model ini transparan dan murah, tetapi hanya menangkap tren, musiman, dan autokorelasi yang bersifat linier. Generasi kedua adalah *machine learning* yang menerima banyak peubah penjelas sekaligus, seperti harga, promosi, hari libur, dan cuaca, serta memodelkan hubungan non-linier tanpa bentuk fungsi yang ditentukan lebih dahulu. Studi pada distributor bahan bangunan menunjukkan bahwa model prediktif dapat diterapkan langsung pada data permintaan perusahaan untuk mendukung optimasi persediaan (Goulart dkk., 2026).

Generasi ketiga adalah jaringan saraf tiruan (*artificial neural networks*/ANN) dan *deep learning*. Praveen dkk. (2019) menggabungkan peramalan deret waktu berbasis AI dengan

ANN untuk menekan biaya persediaan dan biaya kekurangan stok yang bersumber dari galat peramalan, sedangkan Terrada dkk. (2022) mengembangkan model *deep learning* untuk rantai pasok 4.0. Arsitektur berurutan seperti LSTM kini dipakai untuk peramalan dan pengisian ulang otonom pada logistik bersama (Straka & Kleinová, 2026). Generasi keempat adalah *ensemble learning*, yang menggabungkan beberapa model agar galat masing-masing saling menetralkan dan hasilnya lebih stabil. Seyedan dkk. (2023) memakai *ensemble deep learning* sebagai masukan model persediaan, dan Seyedan (2023) menegaskan arah ini dalam disertasinya tentang analitik prediktif untuk peramalan dan persediaan.

Perkembangan ini tidak berarti model statistik sepenuhnya usang. Tadayonrad dan Ndiaye (2023) mengingatkan bahwa evaluasi peramalan untuk keperluan persediaan perlu memperhitungkan keandalan rantai pasok dan musiman, sehingga ukuran galat generik seperti MAPE saja tidak cukup. Pendekatan hibrida, misalnya yang memanfaatkan teori *chaos* dan optimasi (Karpagavigneswari & Kamali, 2025) atau analitik AI yang dipadukan dengan pengambilan keputusan multikriteria kabur (Sandra dkk., 2026), menunjukkan bahwa keunggulan AI paling nyata ketika ia melengkapi, bukan menggantikan, penalaran manajerial. Keterbatasan yang perlu dicatat adalah ketergantungan pada kualitas dan volume data serta rendahnya interpretabilitas model kompleks.

Pengendalian Persediaan Dinamis dan Minimisasi Biaya Operasional

Akurasi peramalan baru bernilai ekonomis ketika diterjemahkan menjadi parameter persediaan. Tiga parameter klasik adalah *safety stock* (SS), titik pemesanan ulang atau *reorder point* (ROP), dan tingkat pesanan maksimum pada kebijakan *order-up-to-level* (S). Dengan galat peramalan berdistribusi mendekati normal, lead time L, periode tinjauan R, dan faktor layanan z yang ditentukan oleh *service level* target, hubungan dasarnya adalah:

$$SS = z \times \sigma_e \times \sqrt{L} \quad (1)$$

$$ROP = \hat{d} \times L + SS \quad (2)$$

$$S = \hat{d} \times (L + R) + z \times \sigma_e \times \sqrt{(L + R)} \quad (3)$$

Dalam persamaan tersebut $\hat{d}$ adalah ramalan permintaan per periode dan σ_e adalah deviasi standar galat peramalan, yang pada praktiknya diaproksimasi dengan RMSE. Pada kebijakan konvensional, σ_e diganti deviasi standar historis yang statis. Pada kebijakan dinamis, $\hat{d}$ dan σ_e diperbarui setiap kali model prediktif dilatih ulang, sehingga SS, ROP, dan S mengikuti kondisi permintaan terkini. Seyedan dkk. (2023) mengilustrasikan pendekatan ini melalui model *order-up-to-level* yang memakai peramalan *ensemble deep learning*.

Di luar formula analitik, kebijakan persediaan juga dapat dipelajari langsung dari interaksi dengan lingkungan. Preil dan Krapp (2022) menerapkan pencarian pohon Monte Carlo untuk manajemen persediaan berbasis AI, dan Yang dkk. (2025) mengusulkan *multi-agent deep reinforcement learning* yang memadukan peramalan permintaan dan optimasi persediaan pada rantai pasok ritel berbasis sensor. Pada filosofi *Just-In-Time* (JIT), peramalan yang lebih akurat dan pembaruan data yang lebih cepat memungkinkan pengiriman dalam lot kecil dan lebih sering tanpa menaikkan risiko kehabisan stok. Aspek real-time ini terutama relevan pada rantai pasok barang mudah rusak (Rao, 2026) dan pada pengisian ulang otonom (Straka & Kleinová, 2026).

Integrasi Sistem Operasional: ERP, IoT, dan *Smart Warehousing*

Model AI hanya sebaik data yang mengalirinya. Sistem *enterprise resource planning* (ERP) menjadi tulang punggung data transaksi, pembelian, dan stok; kajian sistematis atas dampak analitik prediktif berbasis AI dalam ERP menempatkan peramalan permintaan sebagai area yang paling terdampak (Fathima dkk., 2024), dan arsitektur integrasi ERP dengan model AI telah diusulkan untuk optimasi kontrol persediaan (Kumalasari dkk., 2026). Sensor *Internet of Things* (IoT) dan RFID menambah lapisan visibilitas: Jaouhari dkk. (2022) memadukan peramalan berbasis regresi dengan sistem manajemen persediaan berbasis IoT pada perusahaan semikonduktor, sedangkan Ugbebor dkk. (2024) mengulas sistem persediaan otomatis berbasis

IoT yang menurunkan biaya simpan bagi UMKM.

Smart warehousing menyatukan ketiganya: data stok dan pergerakan barang diperbarui secara kontinu, model AI membaca data itu, dan keputusan pengisian ulang diteruskan kembali ke sistem operasional. Penelitian pada rantai pasok ritel berbasis sensor (Yang dkk., 2025) dan pada logistik bersama (Straka & Kleinová, 2026) memperlihatkan bahwa integrasi ini memungkinkan peralihan dari manajemen persediaan yang reaktif ke yang prediktif. Dengan kata lain, integrasi sistem bukan pelengkap, melainkan prasyarat agar akurasi prediktif benar-benar sampai pada keputusan persediaan secara *real-time*.

METODE

Penelitian ini merupakan studi konseptual (*conceptual paper*) dengan desain sintesis literatur sistematis (*conceptual literature synthesis*). Pendekatan ini dipilih karena tujuan artikel adalah membangun kerangka teoretis, bukan menguji hipotesis dengan data primer. Kerangka dibangun secara deduktif: kategori analisis ditetapkan lebih dahulu dari teori manajemen operasional, kemudian literatur dibandingkan pada setiap kategori untuk menemukan pola, kesenjangan, dan hubungan antarkonstruk.

Data bersifat sekunder. Penelusuran literatur dilakukan dengan perangkat lunak *Publish or Perish* pada basis data Google Scholar untuk publikasi tahun 2019–2026 dengan kata kunci ("Artificial Intelligence" AND "Demand Forecasting" AND "Inventory Management"). Hasil penelusuran diekspor dalam format RIS dan memuat 199 rekaman. Kurasi dilakukan bertahap: (1) penghapusan rekaman ganda; (2) penyaringan judul dan abstrak berdasarkan keterkaitan dengan peramalan permintaan atau keputusan persediaan berbasis AI; (3) pemeriksaan kelengkapan metadata dan ketersediaan abstrak; serta (4) penilaian kesesuaian dengan tiga tema analisis, yaitu metode peramalan, pengendalian persediaan dinamis, dan integrasi sistem. Tahap ini menghasilkan 28 artikel rujukan utama. Artikel dipertahankan jika membahas AI atau *machine learning* untuk peramalan atau persediaan serta mengaitkannya dengan biaya, tingkat layanan, atau efisiensi operasional, dan dikeluarkan jika hanya menyinggung topik tersebut sebagai latar tanpa analisis. Rujukan klasik di luar korpus (Box & Jenkins, 1970; Harris, 1913; Lee dkk., 1997) digunakan hanya sebagai landasan konsep.

Ke-28 artikel dianalisis secara deduktif-komparatif: setiap artikel dikodekan menurut metode, tujuan keputusan, konteks industri, dan kontribusinya terhadap biaya atau tingkat layanan, kemudian dibandingkan lintas klaster. Tabel 1 meringkas pemetaan literatur utama tersebut.

Tabel 1 Pemetaan literatur utama (n = 28)

Klaster tematik	Artikel rujukan	n	Kontribusi analitis
A. Tinjauan sistematis dan kritis	Ünal dkk. (2023); Douaioui dkk. (2024); Li dkk. (2025); Hierro dkk. (2026); Walter dkk. (2025); Atwani dkk. (2024); Zhu dkk. (2019)	7	Peta metode, tugas keputusan, dan celah penelitian AI pada peramalan dan persediaan
B. Peramalan berbasis ML/DL dan ensemble	Aktepe dkk. (2021); Praveen dkk. (2019); Terrada dkk. (2022); Goulart dkk. (2026); Tadayonrad & Ndiaye (2023); Karpagavigneswari & Kamali (2025); Sukolkit dkk. (2024); Seyedan (2023)	8	Perbandingan metode, ukuran kinerja peramalan, dan kaitan galat dengan biaya persediaan

Klaster tematik	Artikel rujukan	n	Kontribusi analitis
C. Optimasi dan kebijakan persediaan dinamis	Seyedan dkk. (2023); Preil & Krapp (2022); Yang dkk. (2025); Sandra dkk. (2026); Tang dkk. (2023)	5	Penerjemahan ramalan menjadi kebijakan: order-up-to-level, pencarian pohon Monte Carlo, pembelajaran penguatan
D. Integrasi ERP, IoT, dan pergudangan cerdas	Kumalasari dkk. (2026); Fathima dkk. (2024); Jaouhari dkk. (2022); Ugbebor dkk. (2024); Straka & Kleinová (2026); Rao (2026)	6	Prasyarat data dan arsitektur untuk keputusan persediaan real-time, termasuk barang mudah rusak
E. Konteks UMKM dan Indonesia	Khawarizmi dkk. (2026); Purba dkk. (2026)	2	Kesiapan adopsi dan hambatan pada UMKM dan pelaku usaha Indonesia

Sumber: Hasil sintesis literatur penulis (2026)

HASIL DAN PEMBAHASAN

Komparasi Paradigma Konvensional vs. *AI-Driven Operations*

Perbandingan pada Tabel 2 menunjukkan bahwa perbedaan kedua paradigma tidak terletak pada satu teknik, melainkan pada cara informasi mengalir dari data ke keputusan. Pada paradigma konvensional, peramalan, penetapan parameter, dan pemesanan adalah tiga langkah terpisah yang dijalankan berkala dengan asumsi statis. Pada paradigma berbasis AI, ketiganya menjadi satu lingkar pembelajaran: ramalan memperbarui parameter, hasil keputusan menghasilkan data baru, dan data baru melatih ulang model (Ünal dkk., 2023; Li dkk., 2025).

Tabel 2 Perbandingan paradigma konvensional dan paradigma berbasis AI

Aspek	Paradigma konvensional	Paradigma AI-driven
Metode peramalan	Moving average, exponential smoothing, ARIMA; univariat, linier, dikalibrasi berkala	ML, ANN, deep learning, dan ensemble; multivariat, non-linier, dilatih ulang secara berkelanjutan
Penentuan safety stock dan ROP	Deviasi standar historis yang statis; EOQ dengan parameter tetap; ditinjau berkala	σ_e dan ramalan diperbarui dinamis; parameter disesuaikan per SKU, lokasi, dan musim; dapat memakai kebijakan hasil pembelajaran penguatan
Mitigasi bullwhip effect	Setiap pihak meramal sendiri; pesanan dalam lot besar; distorsi informasi diteruskan ke hulu	Ramalan berbasis data sinyal permintaan nyata (demand sensing); visibilitas bersama; pesanan lebih halus dan lebih sering
Dampak terhadap biaya simpan	Buffer besar sebagai pelindung ketidakpastian; modal kerja tertahan; risiko keusangan tinggi	Buffer proporsional terhadap ketidakpastian aktual; modal kerja dan risiko keusangan menurun

Aspek	Paradigma konvensional	Paradigma AI-driven
Dampak terhadap biaya kekurangan stok	Kehabisan stok tidak terduga saat promosi atau lonjakan; bergantung pada penilaian manual	Antisipasi lonjakan lebih dini; service level lebih terjaga pada biaya simpan yang sama atau lebih rendah

Sumber: Hasil analisis komparatif penulis (2026), diolah dari klaster A–D pada Tabel 1

Dua catatan perlu ditambahkan. Pertama, keunggulan paradigma AI bersifat bersyarat. Ia bergantung pada ketersediaan data, kualitas pembersihan data, dan kemampuan organisasi memelihara model. Kedua, mitigasi efek cambuk membutuhkan lebih dari peramalan yang akurat di satu titik. Seperti ditunjukkan Lee dkk. (1997), distorsi tumbuh dari pembaruan ramalan yang tidak terkoordinasi antarpihak, sehingga manfaat terbesar baru muncul bila sinyal permintaan hilir dibagikan dan dipakai bersama di sepanjang rantai.

Mekanisme Reduksi Biaya Persediaan melalui Akurasi Prediktif

Hubungan antara akurasi dan biaya dapat dijelaskan melalui Persamaan (1) sampai (3). Karena SS berbanding lurus dengan σ_e, penurunan galat peramalan, yang tercermin pada turunnya RMSE, menurunkan *safety stock* yang dibutuhkan untuk *service level* yang sama. Biaya simpan yang terkait dengan penyangga ini berubah sebesar:

$$\Delta BS = h \times z \times \sqrt{L} \times \Delta\sigma_e \qquad (4)$$

dengan h sebagai biaya simpan per unit per periode. Sebagai ilustrasi hipotetis, bila RMSE turun dari 100 menjadi 70 unit per minggu pada L = 4 minggu dan z = 1,65 (sekitar 95 persen), maka SS turun dari 330 menjadi 231 unit atau sekitar 30 persen. Angka ini bukan temuan empiris, melainkan penegasan bahwa penurunan biaya penyangga sebanding dengan penurunan σ_e. Praveen dkk. (2019) menekankan kaitan serupa: biaya persediaan tinggi dan biaya kekurangan stok yang tidak perlu bersumber dari galat peramalan.

Mekanisme pertama adalah pelepasan modal kerja tertahan. Persediaan rata-rata kira-kira Q/2 + SS, sehingga setiap unit SS yang dapat dihilangkan membebaskan nilai persediaan yang semula membebani kas. Mekanisme kedua adalah pencegahan keusangan dan kerusakan pada barang mudah rusak. Semakin lama barang berada di gudang, semakin besar peluang kedaluwarsa, sehingga pengurangan stok berlebih menurunkan kerugian akibat pembuangan; pengelolaan persediaan *real-time* pada rantai pasok barang mudah rusak merupakan salah satu penerapan yang paling menonjol (Rao, 2026). Mekanisme ketiga adalah peningkatan *service level*: dengan SS yang sama, ramalan yang lebih akurat memperbesar z yang dapat dicapai, atau dengan z yang sama SS yang diperlukan lebih kecil.

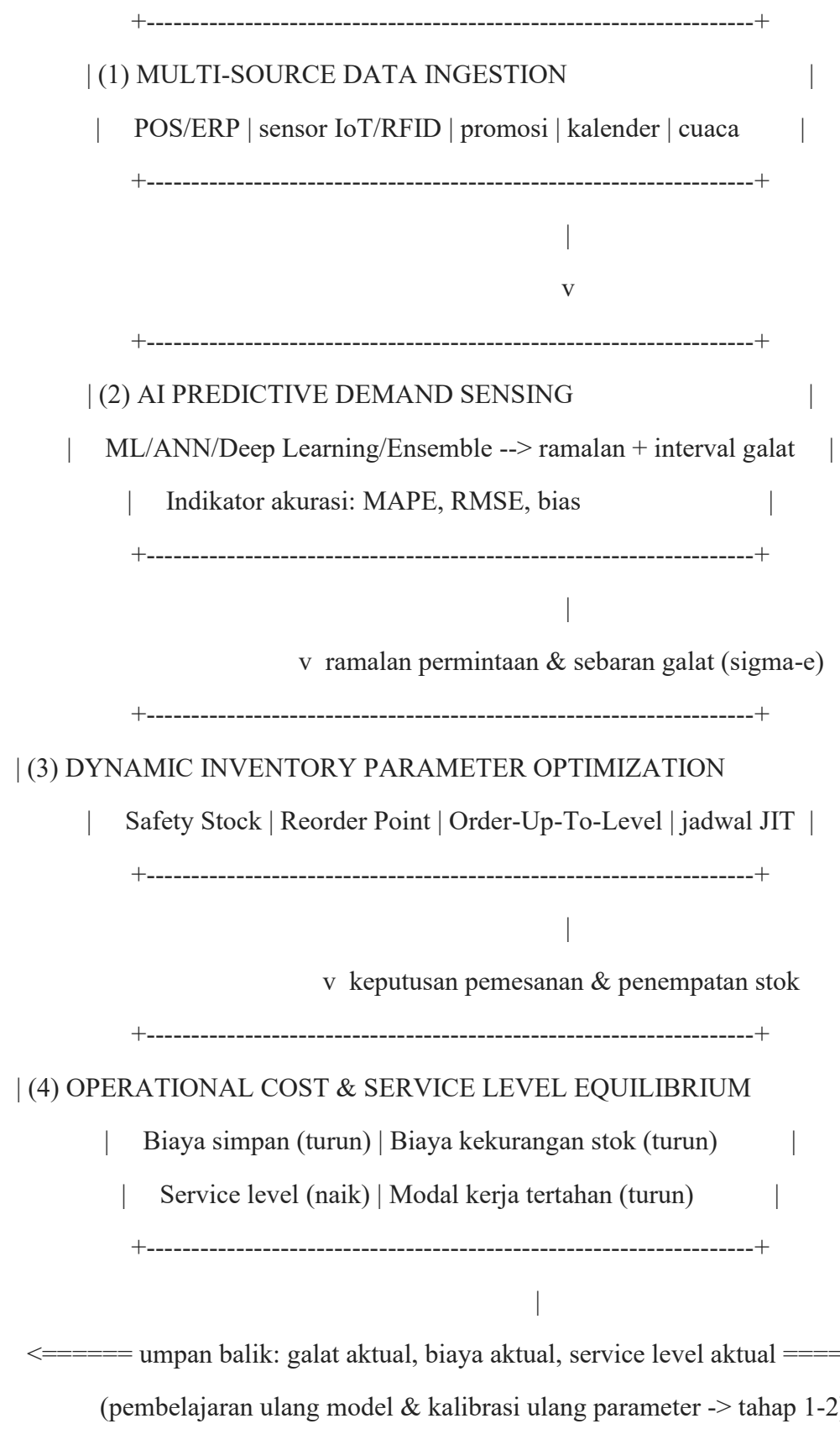
Namun hubungan ini tidak mekanis. Biaya kekurangan stok dan biaya kelebihan stok biasanya tidak simetris. Pada pendekatan *newsvendor*, tingkat persediaan optimal memenuhi $F(S^*) = C_u/(C_u + C_o)$ (5), dengan C_u biaya kekurangan per unit dan C_o biaya kelebihan per unit. Karena itu yang bernilai bukan sekadar MAPE atau RMSE yang kecil, melainkan ramalan yang tidak bias dan menggambarkan ekor sebaran dengan baik. Hal ini sejalan dengan argumen bahwa indikator kinerja peramalan perlu mempertimbangkan keandalan rantai pasok dan musiman (Tadayonrad & Ndiaye, 2023). Berdasarkan uraian ini diajukan dua proposisi:

P1: Semakin rendah galat peramalan (RMSE) yang dihasilkan model AI, semakin rendah *safety stock* yang diperlukan untuk *service level* tertentu, sehingga semakin rendah biaya simpan dan modal kerja tertahan.

P2: Pengaruh akurasi prediktif terhadap biaya total diperkuat oleh asimetri biaya kekurangan dan kelebihan stok, serta oleh kecepatan siklus penyesuaian parameter; pengaruh ini paling besar pada barang mudah rusak dan permintaan yang berfluktuasi.

Proposisi Kerangka Konseptual *AI-Driven Demand & Inventory Optimization* (AIDIO)

Kerangka AIDIO menggabungkan temuan klaster A–E menjadi satu alur empat tahap yang digambarkan pada Gambar 1. Kerangka ini dirancang sebagai lingkar tertutup: keluaran tahap terakhir kembali menjadi data pelatihan bagi tahap awal.



Gambar 1 Kerangka konseptual AIDIO

Sumber: Konstruksi konseptual penulis (2026), diadaptasi dari sintesis klaster A–E

Tahap 1, *Multi-Source Data Ingestion*, mengumpulkan data transaksi dari ERP dan sistem kasir (*point of sale*), data sensor IoT dan RFID tentang posisi stok, serta data eksternal seperti kalender promosi, hari libur, dan cuaca. Integrasi sumber-sumber ini menjadi prasyarat teknis yang ditegaskan oleh kajian ERP (Fathima dkk., 2024; Kumalasari dkk., 2026) dan IoT (Jaouhari dkk., 2022; Ugbebor dkk., 2024). Kualitas data pada tahap ini, mencakup kelengkapan, keseragaman SKU, dan ketepatan waktu, membatasi mutu seluruh tahap berikutnya.

Tahap 2, *AI Predictive Demand Sensing*, menjalankan model ML, ANN, *deep learning*, atau *ensemble* untuk menghasilkan ramalan sekaligus ukuran ketidakpastiannya (σ_e). Pemilihan model mengikuti karakter permintaan: model sederhana cukup untuk permintaan stabil, sedangkan model sekuensial atau *ensemble* lebih sesuai untuk permintaan berpola kompleks (Seyedan dkk., 2023; Straka & Kleinová, 2026). Indikator akurasi (MAPE, RMSE, dan bias) dipantau secara kontinu sesuai anjuran Tadayonrad dan Ndiaye (2023).

Tahap 3, *Dynamic Inventory Parameter Optimization*, menerjemahkan ramalan dan σ_e menjadi *safety stock*, *reorder point*, *order-up-to-level*, dan jadwal pengisian ulang bergaya JIT melalui Persamaan (1) sampai (3), atau melalui kebijakan hasil pembelajaran penguatan bila lingkungan keputusannya kompleks (Preil & Krapp, 2022; Yang dkk., 2025). Parameter dihitung per SKU dan lokasi serta diperbarui mengikuti siklus pelatihan ulang model.

Tahap 4, *Operational Cost & Service Level Equilibrium*, mengevaluasi hasil keputusan melalui empat ukuran: biaya simpan, biaya kekurangan stok, modal kerja tertahan, dan *service level*. Titik keseimbangan tidak ditetapkan sekali, melainkan dicari terus-menerus; bila salah satu ukuran memburuk, galat dan biaya aktual dikembalikan ke Tahap 1 dan 2 untuk pelatihan ulang. Proposisi tambahan yang dihasilkan kerangka ini adalah:

P3: Integrasi data multisumber (ERP, IoT, dan sumber eksternal) memoderasi hubungan antara kecanggihan model dan akurasi peramalan; model kompleks tanpa data terintegrasi tidak menghasilkan akurasi yang lebih tinggi.

P4: Siklus umpan balik yang pendek dari Tahap 4 ke Tahap 1–2 mempertahankan akurasi dan keseimbangan biaya dari waktu ke waktu, sedangkan ketiadaan umpan balik menyebabkan degradasi model dan parameter yang usang.

Implikasi Manajerial dan Konteks Bisnis di Indonesia

Bagi korporasi ritel, implikasi pertamanya adalah memperlakukan peramalan sebagai kapabilitas inti, bukan fungsi pendukung. Ritel modern memiliki data transaksi harian dalam jumlah besar dan sistem ERP yang relatif matang, sehingga lebih siap memulai dari Tahap 1–3 AIDIO pada kategori yang paling berisiko: barang mudah rusak, barang berpromosi tinggi, dan SKU dengan lead time panjang. Karakter kepulauan Indonesia membuat lead time pasokan beragam antarwilayah; variabilitas ini langsung memperbesar $\sqrt{L}$ dan σ_e pada Persamaan (1), sehingga manfaat parameter dinamis diperkirakan paling besar pada jaringan distribusi yang melintasi banyak pulau. Selain itu, manajer perlu menetapkan ukuran keberhasilan berbasis biaya dan *service level*, bukan hanya MAPE, serta menyiapkan tata kelola data dan keterampilan analitik di dalam organisasi.

Bagi UMKM, hambatan utama bukan algoritma, melainkan data, biaya, dan keterampilan. Kajian sistematis tentang AI untuk peramalan permintaan pada UMKM tekstil Indonesia (Khawarizmi dkk., 2026) dan studi kualitatif pelaku usaha (Purba dkk., 2026) menunjukkan bahwa pemanfaatan AI dalam peramalan masih berada pada tahap awal. Sistem persediaan otomatis berbasis IoT juga dinilai dapat menekan biaya simpan UMKM, tetapi investasi awalnya menjadi pertimbangan (Ugbebor dkk., 2024). Oleh karena itu adopsi sebaiknya bertahap: (i) digitalisasi pencatatan stok dan penjualan harian dengan peramalan sederhana; (ii) peralihan ke model *machine learning* ringan atau layanan komputasi awan berbiaya rendah; dan (iii) integrasi ERP atau IoT hanya setelah manfaat tahap sebelumnya terbukti.

Pendekatan bertahap ini menjaga agar investasi tetap sebanding dengan penurunan biaya yang dapat diharapkan.

Implikasi lain adalah perlunya kewaspadaan terhadap risiko model. Model dapat mengalami penurunan kinerja ketika pola permintaan berubah (*model drift*), sementara model yang tidak dapat dijelaskan sulit dipercaya oleh pengambil keputusan. Karena itu keputusan akhir pada barang bernilai tinggi atau berisiko tinggi sebaiknya tetap melibatkan penilaian manusia (*human-in-the-loop*), dan kinerja model perlu diaudit secara berkala.

SIMPULAN

Artikel ini menyintesis 28 artikel terpilih dari penelusuran Google Scholar periode 2019–2026 untuk menunjukkan bahwa peramalan permintaan dan pengendalian persediaan paling tepat dipandang sebagai satu sistem keputusan, bukan dua fungsi yang terpisah. Kontribusi teoretisnya ada tiga. Pertama, artikel ini menghubungkan evolusi metode peramalan dari model statistik linier ke ML, ANN, *deep learning*, dan *ensemble* dengan parameter persediaan dinamis melalui hubungan eksplisit antara galat peramalan, *safety stock*, dan biaya. Kedua, artikel ini menjelaskan tiga mekanisme reduksi biaya, yaitu pelepasan modal kerja tertahan, pencegahan keusangan, dan peningkatan *service level*, sekaligus menunjukkan bahwa hubungannya dikondisikan oleh asimetri biaya dan bias peramalan. Ketiga, artikel ini merumuskan kerangka AIDIO dengan empat tahap dan empat proposisi yang dapat diuji.

Secara manajerial, hasil sintesis menyarankan agar korporasi ritel memulai dari kategori berisiko tinggi dengan data terintegrasi, sedangkan UMKM mengadopsi AI secara bertahap sesuai kesiapan data, dana, dan keterampilan, dengan ukuran keberhasilan berupa biaya total dan *service level*. Studi ini memiliki keterbatasan: korpus bersumber dari satu basis data dengan kualitas indeksasi beragam, sebagian artikel berupa prosiding atau prapublikasi, ilustrasi numerik bersifat hipotetis, dan kerangka belum diuji secara empiris. Agenda riset berikutnya mencakup pengujian proposisi P1 sampai P4 melalui studi kasus atau eksperimen terkendali pada ritel dan UMKM Indonesia, penilaian nilai tambah peramalan terhadap biaya total dengan data nyata, serta penelitian lintas rantai pasok untuk mengukur dampak berbagi sinyal permintaan terhadap efek cambuk.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Universitas Muslim Indonesia, khususnya Program Pascasarjana, atas dukungan akademik dan fasilitas yang diberikan selama proses penyusunan artikel ini. Ucapan terima kasih juga disampaikan kepada para dosen dan rekan akademisi yang telah memberikan arahan, masukan, serta dukungan dalam pengembangan gagasan dan penyempurnaan kajian ini. Penulis turut mengapresiasi seluruh peneliti yang karya ilmiahnya menjadi sumber referensi dalam proses sintesis literatur. Diharapkan artikel ini dapat memberikan kontribusi bagi pengembangan ilmu pengetahuan, khususnya dalam bidang manajemen operasional, peramalan permintaan, pengendalian persediaan, dan penerapan kecerdasan buatan untuk meningkatkan efisiensi biaya operasional.

DAFTAR PUSTAKA

Aktepe, A., Yanık, E., & Ersöz, S. (2021). Demand forecasting application with regression and artificial intelligence methods in a construction machinery company. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-021-01737-8

Atwani, M., Hlyal, M., & Alami, J. E. (2024). AI-based demand forecasting models: A systematic

literature review. *International Journal of Industrial Engineering*.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. Holden-Day.

Douaioui, K., Oucheikh, R., Benmoussa, O., dkk. (2024). Machine learning and deep learning models for demand forecasting in supply chain management: A critical review. *Applied System Innovation, 7*(5), 93. https://www.mdpi.com/2571-5577/7/5/93

Fathima, F., Inparaj, R., Thuvarakan, D., dkk. (2024). Impact of AI-based predictive analytics on demand forecasting in ERP systems: A systematic literature review. *Prosiding konferensi IEEE tentang komputasi cerdas*. https://ieeexplore.ieee.org/abstract/document/10550480/

Goulart, D. D., Carvalho, R. B. de, dkk. (2026). Predictive models for inventory optimization: A machine learning application for demand forecasting at a construction supplies distributor. *Future Business Journal*. https://doi.org/10.1186/s43093-026-00807-8

Harris, F. W. (1913). How many parts to make at once. *Factory, The Magazine of Management, 10*(2), 135–136, 152.

Hierro, D. M. B. del, Martínez, D. M. C., dkk. (2026). Intelligent algorithms in inventory management: A systematic literature review. *Algorithms, 19*(9), 711. https://www.mdpi.com/1999-4893/19/9/711

Jaouhari, A. E., Alhilali, Z., Arif, J., Fellaki, S., dkk. (2022). Demand forecasting application with regression and IoT based inventory management system: A case study of a semiconductor manufacturing company. *International Journal of Engineering Research in Africa, 60*, 189. https://www.scientific.net/JERA.60.189

Karpagavigneswari, T., & Kamali, R. (2025). Intelligent inventory demand forecasting: Using chaos theory and optimization for supply chain resilience. *OPSEARCH*. https://doi.org/10.1007/s12597-025-01005-w

Khawarizmi, I. N. A., Gunawan, G., Parung, J., dkk. (2026). Artificial intelligence in demand forecasting for textile MSMEs supply chains in Indonesia: A systematic literature review. *Vokatek*. https://journal.diginus.id/VOKATEK/article/view/1272

Kumalasari, K., Pane, E. P., dkk. (2026). ERP system integration architecture with artificial intelligence models for supply chain management and inventory control optimization. *JAISEN*. http://journal.unilak.ac.id/index.php/jaisen/article/view/32555

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). The bullwhip effect in supply chains. *Sloan Management Review, 38*(3), 93–102.

Li, J., Huang, H., Gong, Y., Wang, L., dkk. (2025). A review of artificial intelligence in inventory management: Methods, applications and directions. https://pdfs.semanticscholar.org/03f3/18203419bed30111c1075b1f76e866004c72.pdf

Praveen, U., Farnaz, G., & Hatim, G. (2019). Inventory management and cost reduction of supply chain processes using AI based time-series forecasting and ANN modeling. *Procedia Manufacturing*. https://www.sciencedirect.com/science/article/pii/S2351978920300354

Preil, D., & Krapp, M. (2022). Artificial intelligence-based inventory management: A Monte Carlo tree search approach. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-03935-2

Purba, P., Siahaan, E. M., Hulu, C. W., dkk. (2026). Utilization of artificial intelligence in demand forecasting: A qualitative study of business actors. *Journal of Social Science*. https://journal.taromboresearch.com/index.php/joss/article/view/44

Rao, V. V. (2026). Leveraging AI for real-time inventory management in perishable goods supply chains. Dalam *Modernizing the food industry: AI-powered solutions*. IGI Global. https://www.igi-global.com/chapter/leveraging-ai-for-real-time-inventory-management-in-perishable-goods-supply-chains/385274

Sandra, M., Narayanamoorthy, S., Suvitha, K., dkk. (2026). A novel AI analytics driven forecasting

paradigm for strategic inventory management through fuzzy multi-criteria decision model in supply chains. *International Journal of Fuzzy Systems*. https://doi.org/10.1007/s40815-025-02007-y

Seyedan, M., Mafakheri, F., & Wang, C. (2023). Order-up-to-level inventory optimization model using time-series demand forecasting with ensemble deep learning. *Supply Chain Analytics*. https://www.sciencedirect.com/science/article/pii/S2949863523000237

Seyedan, S. (2023). *Development of predictive analytics for demand forecasting and inventory management in supply chain using machine learning techniques* [Disertasi doktoral, Concordia University]. https://spectrum.library.concordia.ca/id/eprint/993482/

Straka, M., & Kleinová, K. (2026). Deep learning-enhanced proactive strategy: LSTM and VRP/ACO for autonomous replenishment and demand forecasting in shared logistics. *Applied Sciences, 16*(6), 2838. https://www.mdpi.com/2076-3417/16/6/2838

Sukolkit, N., Arunyanart, S., & Apichottanakul, A. (2024). An open innovative inventory management based demand forecasting approach for the steel industry. *Journal of Open Innovation: Technology, Market, and Complexity*. https://www.sciencedirect.com/science/article/pii/S2199853124002014

Tadayonrad, Y., & Ndiaye, A. B. (2023). A new key performance indicator model for demand forecasting in inventory management considering supply chain reliability and seasonality. *Supply Chain Analytics*. https://www.sciencedirect.com/science/article/pii/S2949863523000250

Tang, Y. M., Chau, K. Y., Lau, Y., & Zheng, Z. (2023). Data-intensive inventory forecasting with artificial intelligence models for cross-border e-commerce service automation. *Applied Sciences, 13*(5), 3051. https://www.mdpi.com/2076-3417/13/5/3051

Terrada, L., Khaili, M. E., & Ouajji, H. (2022). Demand forecasting model using deep learning methods for supply chain management 4.0. *International Journal of Advanced Computer Science and Applications*.

Ugbebor, F., Adeteye, M., & Ugbebor, J. (2024). Automated inventory management systems with IoT integration to optimize stock levels and reduce carrying costs for SMEs: A comprehensive review. *Journal of Artificial Intelligence General Science, 6*(1), 306–340. https://ideas.repec.org/a/das/njaigs/v6y2024i1p306-340id257.html

Ünal, Ö. A., Erkayman, B., & Usanmaz, B. (2023). Applications of artificial intelligence in inventory management: A systematic review of the literature. *Archives of Computational Methods in Engineering*. https://doi.org/10.1007/s11831-022-09879-5

Walter, A., Ahsan, K., & Rahman, S. (2025). Application of artificial intelligence in demand planning for supply chains: A systematic literature review. *The International Journal of Logistics Management, 36*(3), 672. https://www.emerald.com/ijlm/article/36/3/672/1246964

Yang, Y., Wang, M., Wang, J., Li, P., & Zhou, M. (2025). Multi-agent deep reinforcement learning for integrated demand forecasting and inventory optimization in sensor-enabled retail supply chains. *Sensors, 25*(8), 2428. https://www.mdpi.com/1424-8220/25/8/2428

Zhu, X., Zhang, G., & Sun, B. (2019). A comprehensive literature review of the demand forecasting methods of emergency resources from the perspective of artificial intelligence. *Natural Hazards*. https://doi.org/10.1007/s11069-019-03626-z