



# Implementasi Sistem Informasi Manufaktur Berbasis Web dengan Pendekatan Hybrid Rule-Based dan Machine Learning untuk Evaluasi Kinerja Pemasok pada Industri Otomotif

Dody Mulyadi, Cahyono Budy Santoso\*

Fakultas Teknologi dan Desain, Program Studi Sistem Informasi, Universitas Pembangunan Jaya, Tangerang Selatan, Indonesia

Email: <sup>1</sup>dody.mulyad@student.upj.ac.id, <sup>2,\*</sup>cahyono.budy@upj.ac.id

Email Penulis Korespondensi: cahyono.budy@upj.ac.id

**Abstrak**—Evaluasi kinerja pemasok (supplier performance evaluation) pada industri manufaktur otomotif merupakan aktivitas kritis yang menentukan kontinuitas lini produksi. Namun proses ini masih dominan dilakukan secara manual sehingga mengakibatkan inefisiensi administratif, inkonsistensi data, dan lambatnya pengambilan keputusan. Penelitian ini bertujuan merancang dan mengimplementasikan sistem informasi manufaktur berbasis web yang mengintegrasikan pendekatan hybrid rule-based dan machine learning untuk mengoptimalkan evaluasi kinerja pemasok di PT ABC. Dataset mencakup 1.008 catatan transaksi dari 28 pemasok selama tiga tahun (2022–2024) dengan tujuh kriteria evaluasi yaitu Accident, Incident, Line Stop, Off Line, Kanban Delay, Laporan Masalah Delivery (LMD), dan Delay Delivery. Metode penelitian menggunakan Research and Development (R&D) dengan model SDLC Waterfall yang dipertajam metodologi CRISP-DM untuk komponen analitik. Rekayasa fitur (feature engineering) menghasilkan 22 variabel input melalui teknik lag-1, trend analysis, dan rolling average, sementara ketidakseimbangan kelas diatasi dengan Synthetic Minority Over-sampling Technique (SMOTE). Tiga algoritma ensemble (Random Forest, XGBoost, dan Gradient Boosting) dievaluasi melalui 5-Fold Stratified Cross Validation. XGBoost terpilih sebagai model terbaik dengan akurasi 88,82% dan F1-Score Macro 88,80%. Komponen hybrid fusion layer berhasil menghasilkan rekomendasi tindakan bertingkat dalam lima kategori urgensi, dengan akurasi prediksi pada data operasional aktual mencapai 93,16%. Kontribusi penelitian ini bagi perkembangan ilmu pengetahuan adalah integrasi konsep decision support system berbasis kecerdasan buatan dengan platform sistem informasi manufaktur operasional, sekaligus menyediakan kerangka kerja hybrid yang dapat direplikasi pada konteks industri manufaktur lain di Indonesia.

**Kata Kunci:** Evaluasi Kinerja Pemasok; Machine Learning; XGBoost; SMOTE; Hybrid Rule-Based

**Abstract**—Supplier performance evaluation in the automotive manufacturing industry is a critical activity that determines production line continuity. However, this process remains predominantly manual, resulting in administrative inefficiency, data inconsistency, and slow decision-making. This study aims to design and implement a web-based manufacturing information system that integrates a hybrid rule-based and machine learning approach to optimize supplier performance evaluation at PT ABC. The dataset comprises 1,008 transaction records from 28 suppliers over three years (2022–2024) with seven evaluation criteria: Accident, Incident, Line Stop, Off Line, Kanban Delay, Delivery Problem Report (LMD), and Delay Delivery. The research methodology employs Research and Development (R&D) with the Waterfall SDLC model enriched by the CRISP-DM methodology for the analytical component. Feature engineering produced 22 input variables through lag-1, trend analysis, and rolling average techniques, while class imbalance was addressed using SMOTE. Three ensemble algorithms (Random Forest, XGBoost, and Gradient Boosting) were evaluated through 5-Fold Stratified Cross Validation. XGBoost was selected as the best model with 88.82% accuracy and 88.80% Macro F1-Score. The hybrid fusion layer successfully generated tiered action recommendations across five urgency categories, with prediction accuracy on actual operational data reaching 93.16%. The contribution of this research to the development of scientific knowledge is the integration of an AI-based decision support system concept with an operational manufacturing information system platform, while providing a replicable hybrid framework for other manufacturing industry contexts in Indonesia.

**Keywords:** Supplier Performance Evaluation; Machine Learning; XGBoost; SMOTE; Hybrid Rule-Based

## 1. PENDAHULUAN

Industri manufaktur global tengah mengalami pergeseran paradigma fundamental akibat gelombang transformasi digital dan implementasi Industry 4.0. Dalam konteks ini, adopsi teknologi berbasis digital supply chain telah menjadi prasyarat esensial bagi keberlanjutan operasi perusahaan yang ingin mempertahankan daya saing (Jing & Fan, 2024). Transformasi digital secara empiris terbukti memberikan manfaat nyata berupa peningkatan visibilitas informasi, kemampuan analisis real-time, serta efisiensi fungsi supply chain secara keseluruhan (He et al., 2024). Kemampuan adaptif perusahaan untuk mengintegrasikan sistem informasi modern ke dalam manajemen hubungan pemasok (supplier relationship management) diakui sebagai faktor penentu penciptaan keunggulan kompetitif jangka panjang (Smith et al., 2024). Dalam ekosistem manufaktur otomotif yang menuntut presisi dan kontinuitas tinggi, integrasi sistem informasi tersebut bukan lagi pilihan strategis melainkan kebutuhan operasional yang harus dipenuhi.

Evaluasi kinerja pemasok merupakan pilar kritis dalam sistem manajemen rantai pasok yang efektif. Kualitas produk akhir dan kelancaran lini produksi sangat bergantung pada kinerja jaringan pemasok. Kinerja yang berada di bawah standar tercermin dari indikator seperti ketidaksesuaian kualitas, insiden cacat produk, keterlambatan pengiriman, atau masalah administrasi yang berdampak langsung pada terhentinya proses produksi (line stop, offline) dan pembengkakan biaya overhead (Govindan et al., 2023). Oleh karena itu pemantauan kinerja pemasok harus dilakukan secara sistematis, terukur, dan berkesinambungan. Riset terkini menunjukkan bahwa pemanfaatan data digital untuk evaluasi pemasok mampu memitigasi risiko disrupti operasional secara signifikan (Santoso & Widodo, 2021).



Kondisi inilah yang mendorong perlunya pengembangan sistem evaluasi pemasok berbasis teknologi cerdas yang mampu menyajikan analisis multi-kriteria secara otomatis dan responsif.

Berbagai penelitian terdahulu telah menyoroti pentingnya integrasi sistem informasi dalam supply chain management (SCM), namun pendekatannya bervariasi dan masing-masing memiliki keterbatasan. Pendekatan pertama berbasis sistem informasi manajemen inventori konvensional: Nasution et al. (2022) membuktikan peningkatan akurasi stok melalui aplikasi berbasis web, Ramdhani dan Supena (2022) menekan selisih pencatatan inventori melalui sistem manajemen persediaan, dan Supriyatna et al. (2023) mengoptimalkan proses pembelian bahan baku dengan platform web. Pendekatan ini efektif untuk pemantauan stok fisik namun tidak menyentuh aspek penilaian kinerja pemasok berbasis indikator operasional yang kompleks. Pendekatan kedua berbasis Multi-Criteria Decision Making (MCDM) seperti AHP, TOPSIS, dan Best-Worst Method: Govindan et al. (2023) dan Chung et al. (2023) mengembangkan kerangka MCDM untuk seleksi pemasok berkelanjutan. Keunggulan MCDM adalah transparansi bobot kriteria, namun bersifat statis, sangat bergantung pada penilaian pakar, dan kurang adaptif terhadap perubahan pola kinerja pemasok dari waktu ke waktu. Pendekatan ketiga berbasis machine learning murni: Zhou et al. (2024) mendemonstrasikan keunggulan algoritma ensemble seperti Random Forest dan XGBoost untuk prediksi kualitas manufaktur. Pendekatan ini menawarkan akurasi prediktif tinggi namun cenderung beroperasi sebagai black box yang sulit dijelaskan kepada pengambil keputusan operasional.

Kesenjangan penelitian (research gap) yang teridentifikasi adalah ketiadaan kerangka kerja terintegrasi yang sekaligus (i) menyediakan platform operasional berbasis web yang siap digunakan oleh praktisi di lantai produksi, (ii) memuat scorecard multi-kriteria dengan mekanisme pembobotan yang transparan dan dapat diaudit, serta (iii) menambahkan kemampuan prediktif probabilistik untuk memberikan peringatan dini terhadap potensi degradasi performa pemasok. Mohammadiojdan et al. (2024) baru-baru ini mendemonstrasikan kerangka kerja hybrid MCDM dengan machine learning untuk seleksi pemasok yang lebih akurat dan explainable, namun studi tersebut masih berhenti pada level konseptual dan belum mengimplementasikannya ke dalam sistem informasi operasional yang utuh. Algoritma ensemble seperti Random Forest (Breiman, 2001) dan XGBoost (Chen & Guestrin, 2016) terbukti unggul dalam memprediksi performa di lingkungan industri (Zhou et al., 2024), dan teknik SMOTE (Chawla et al., 2002) telah menjadi standar penanganan ketidakseimbangan kelas pada klasifikasi multi-kelas (Taskiran et al., 2025). Meski demikian, integrasi komprehensif antara rule-based business logic dan prediksi machine learning ke dalam satu platform web evaluasi pemasok manufaktur masih sangat jarang dijumpai dalam literatur, khususnya pada konteks industri otomotif Indonesia.

PT ABC, perusahaan manufaktur otomotif di Indonesia, menghadapi kendala signifikan dalam evaluasi supplier performance eksistingnya. Tim internal harus mengunduh total pesanan bulanan, merekam masalah pengiriman dari Accident, Incident, Line Stop, Offline, Kanban Delay, Laporan Masalah Delivery (LMD), hingga Delay Delivery, lalu mengompilasi data secara manual menggunakan pivot table pada spreadsheet konvensional. Proses ini menimbulkan beban administratif tinggi, rentan terhadap kesalahan manusia, dan mengakibatkan keterlambatan umpan balik kepada pemasok sehingga tindakan korektif tidak dapat dieksekusi secara real-time. Kondisi ini selaras dengan temuan Lee et al. (2024) bahwa kesenjangan digital dalam pengelolaan rantai pasok menjadi hambatan utama efisiensi operasional industri manufaktur.

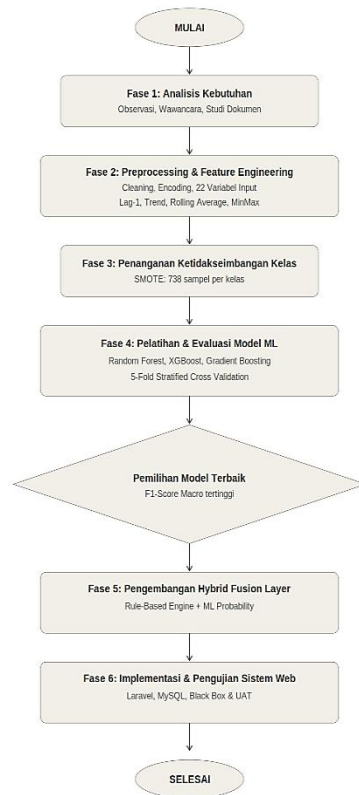
Berdasarkan latar belakang tersebut, penelitian ini bertujuan merancang Sistem Informasi Manufaktur Berbasis Web yang mengintegrasikan pendekatan hybrid rule-based dan machine learning untuk mengoptimalkan evaluasi kinerja pemasok di PT ABC. Pendekatan yang diusulkan dibedakan dari penelitian sebelumnya melalui tiga aspek perbandingan utama. Pertama, dibandingkan studi Nasution et al. (2022), Ramdhani dan Supena (2022), serta Supriyatna et al. (2023) yang fokus pada inventori, penelitian ini menambahkan layer evaluasi scorecard tujuh kriteria dengan pembobotan berbasis moda pengiriman. Kedua, dibandingkan studi MCDM konseptual Govindan et al. (2023), Chung et al. (2023), dan Mohammadiojdan et al. (2024), penelitian ini mengimplementasikan kerangka tersebut ke dalam sistem informasi operasional siap pakai. Ketiga, dibandingkan studi machine learning murni Zhou et al. (2024), penelitian ini menambahkan rule-based engine sebagai lapisan explainability untuk meningkatkan kepercayaan pengguna. Kontribusi utama mencakup: (1) pengembangan hybrid fusion layer yang menggabungkan aturan bisnis deterministik dengan prediksi probabilistik ML untuk menghasilkan rekomendasi tindakan yang lebih akurat dan proaktif; (2) implementasi feature engineering berbasis domain yang mengekstraksi pola temporal dari data historis pemasok; dan (3) validasi empiris pendekatan hybrid yang mengisi kesenjangan antara studi MCDM konseptual dan implementasi sistem informasi operasional di industri manufaktur Indonesia.

## 2. METODOLOGI PENELITIAN

### 2.1 Kerangka Dasar Penelitian

Penelitian ini menggunakan pendekatan Research and Development (R&D) yang berfokus pada perancangan, pengembangan, dan validasi artefak sistem informasi manufaktur berbasis web (Pargaonkar, 2023). Pengembangan sistem menggunakan model SDLC Waterfall yang diperkaya dengan metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining) untuk komponen machine learning. Model Waterfall dipilih karena memberikan kerangka kerja sistematis dan berurutan yang sangat krusial untuk pengembangan sistem business-critical di lingkungan manufaktur (Madihah et al., 2024). Lokasi penelitian dilaksanakan di PT ABC, perusahaan manufaktur otomotif di

Indonesia. Hipotesis utama penelitian adalah bahwa integrasi machine learning ke dalam sistem evaluasi pemasok berbasis web dapat meningkatkan akurasi prediksi kinerja, objektivitas rekomendasi tindakan, dan efisiensi siklus evaluasi dibandingkan metode konvensional. Tahapan penelitian secara keseluruhan diilustrasikan pada Gambar 1 yang menggambarkan alur kerja dari analisis kebutuhan hingga implementasi sistem web.



**Gambar 1.** Diagram Alir Tahapan Penelitian Sistem Rekomendasi Evaluasi Kinerja Pemasok

Sebagaimana ditunjukkan pada Gambar 1, penelitian dijalankan melalui enam fase utama yang saling terkait. Setiap fase memiliki keluaran yang menjadi masukan bagi fase berikutnya, mencerminkan karakter sekuensial Waterfall sekaligus iterasi pemodelan pada CRISP-DM. Titik keputusan utama pada fase pemilihan model terbaik (Fase 4) memastikan bahwa hanya model dengan F1-Score Macro tertinggi yang dilanjutkan ke fase integrasi hybrid fusion layer.

## 2.2 Dataset dan Variabel Penelitian

Dataset yang digunakan mencakup 1.008 catatan transaksi historis dari 28 pemasok aktif selama periode tiga tahun (Januari 2022–Desember 2024), menghasilkan distribusi kelas Rank A: 730 data (75,5%), Rank B: 183 (17,9%), Rank C: 62 (4,5%), Rank D: 24 (1,8%), dan Rank E: 9 (0,4%). Tujuh variabel independen yang berfungsi sebagai kriteria evaluasi adalah Accident, Incident, Line Stop, Off Line, Kanban Delay, Laporan Masalah Delivery (LMD), dan Delay Delivery. Variabel dependen adalah Rank pemasok (A–E) yang dihitung menggunakan Weighted Sum Model (WSM) dengan pembobotan berbeda untuk moda pengiriman Direct dan Milkrun sesuai Tabel 1. Tabel 1 menyajikan kriteria penilaian, skor untuk masing-masing rentang kejadian, dan bobot yang diberikan pada masing-masing moda pengiriman. Bobot Line Stop pada moda Milkrun lebih tinggi (40%) dibandingkan Direct (30%) karena karakteristik konsolidasi pengiriman Milkrun menyebabkan dampak Line Stop bersifat sistemik terhadap beberapa pemasok sekaligus.

**Tabel 1.** Kriteria Penilaian, Skor, dan Pembobotan Evaluasi Supplier Performance PT ABC

No	Criteria	Point	Bobot			
			Direct	Milkrun		
1	Kanban Delay	0	1 - 25%	> 25%	5%	20%
		5	3	1		
2	Accident	0	1 - 2	> 2	20%	
		5	3	1		
3	Incident	0	1 - 2	> 2	20%	
		5	3	1		
4	Line Stop	0	1 - 10	>10	30%	40%
		5	3	1		

No	Criteria	Point	Bobot			
			Direct	Milkrun		
5	Off line	0	1 - 10	>10	10%	30%
		5	3	1		
6	LMD	0	1 - 2	> 2	10%	10%
		5	3	1		
7	Delay Delivery	0	1 - 2	> 2	5%	
		5	3	1		

Tabel 1 menyajikan kerangka scorecard multi-kriteria yang menjadi instrumen utama penilaian kinerja pemasok di PT ABC, dengan mengintegrasikan tiga elemen secara simultan: tujuh kriteria evaluasi, sistem skoring berbasis rentang kejadian, dan pembobotan yang dibedakan menurut moda pengiriman. Sistem skoring menerapkan skala ganjil 5–3–1 di mana skor 5 mencerminkan kondisi tanpa kejadian, skor 3 untuk rentang menengah, dan skor 1 untuk rentang yang melampaui ambang batas. Penggunaan skala ganjil ini bertujuan menghasilkan diferensiasi yang tegas antar kategori kinerja sekaligus menghindari kecenderungan respons netral pada zona menengah.

Diferensiasi bobot antar moda mencerminkan karakteristik logistik kedua skema pengiriman. Pada moda Milkrun, kriteria yang berdampak sistemik terhadap konsolidasi rute memperoleh bobot lebih tinggi yaitu Line Stop (40% vs 30%), Off Line (30% vs 10%), dan Kanban Delay (20% vs 5%), sementara Accident, Incident, dan Delay Delivery tidak diberi bobot pada Milkrun karena tanggung jawab tersebut berpindah ke penyedia jasa logistik konsolidator. Total bobot pada masing-masing moda terverifikasi tepat 100%, sehingga tabel valid digunakan sebagai dasar perhitungan Weighted Sum Model (WSM). Nilai POINT yang dihasilkan kemudian dikonversi menjadi Rank A–E menurut ambang manajemen, dan menjadi ground truth bagi pelatihan model machine learning sekaligus basis aturan dalam rule-based engine pada hybrid fusion layer.

### 2.3 Tahapan Penelitian dan Pipeline Machine Learning

Pengembangan sistem dijalankan melalui enam fase yang saling berkesinambungan:

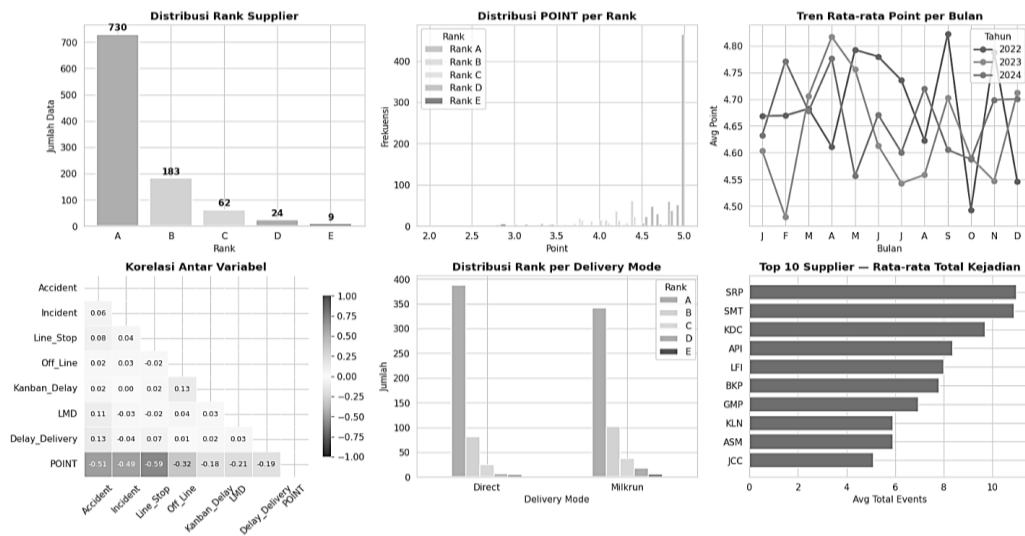
- Fase 1: Analisis Kebutuhan. Pengumpulan data dilakukan melalui observasi langsung terhadap alur kerja pengadaan di PT ABC, wawancara mendalam dengan pengguna kunci, serta studi dokumentasi terhadap historis kinerja pemasok dan SOP yang berlaku (Nasution et al., 2022).
- Fase 2: Preprocessing dan Feature Engineering. Data mentah diproses melalui: (a) pembersihan dan validasi data tidak ditemukan *missing values*; (b) *encoding* variabel kategorikal menggunakan *LabelEncoder* (*Delivery Mode: Direct=0, Milkrun=1; Month: 0–11*); (c) rekayasa fitur menghasilkan 22 variabel input melalui fitur *lag-1* (nilai bulan sebelumnya), *trend* (delta antar periode), *rolling average* 3 bulan, total kejadian kumulatif, *risk flag*, dan *risk score* komposit; serta (d) normalisasi *MinMaxScaler* ke rentang [0,1] agar seluruh fitur berkontribusi setara terhadap model.
- Fase 3: Penanganan Ketidakseimbangan Kelas. Distribusi kelas target yang sangat timpang diatasi menggunakan SMOTE (Chawla et al., 2002). Teknik ini menghasilkan data sintesis untuk kelas minoritas (Rank D dan E) dengan menginterpolasi antara sampel-sampel terdekat (*k-nearest neighbors*), mengubah distribusi dari kondisi awal yang tidak seimbang menjadi 738 sampel per kelas untuk pelatihan model yang lebih adil (Taskiran et al., 2025).
- Fase 4: Pelatihan dan Evaluasi Model ML. Tiga algoritma *ensemble learning* dilatih dan dievaluasi menggunakan *5-Fold Stratified Cross Validation*: (a) *Random Forest* (Breiman, 2001) — *n\_estimators=80, max\_depth=7, class\_weight='balanced'*; (b) *XGBoost* (Chen & Guestrin, 2016) *n\_estimators=80, max\_depth=5, learning\_rate=0,1, eval\_metric='mlogloss'*; (c) *Gradient Boosting* (Friedman, 2001) *n\_estimators=80, max\_depth=4, learning\_rate=0,1*. Metrik evaluasi yang digunakan adalah *Accuracy, F1-Score Macro, Precision Macro, dan Recall Macro*. Target prediksi adalah *Rank* pemasok pada bulan berikutnya, memungkinkan sistem memberikan peringatan dini secara proaktif.
- Fase 5: Pengembangan Hybrid Fusion Layer. Sistem *hybrid* dikembangkan dengan mengintegrasikan: (a) *Rule-based Engine* yang berisi aturan bisnis deterministik berbasis *threshold* operasional PT ABC; dan (b) prediksi probabilistik model ML terbaik. *Fusion layer* menggabungkan tingkat urgensi dari aturan bisnis dengan arah perubahan *rank* yang diprediksi dan *confidence score* ML, menghasilkan lima kategori urgensi: *Critical, High, Medium, Low, dan Improving*.
- Fase 6: Implementasi dan Pengujian Sistem Web. Arsitektur *back-end* dibangun menggunakan Laravel (PHP 8.x) dengan MySQL 8.0. *Front-end* menggunakan Bootstrap 5 dan Chart.js. Pengujian dilaksanakan melalui *Black Box Testing* dan *User Acceptance Testing* (UAT) oleh 18 responden (3 Manajer Pengadaan, 8 Staf Pengadaan, 4 Staf Administrasi, 3 Manajer Produksi) menggunakan kuesioner skala Likert 5 poin (Zulkarnaini et al., 2023).

## 3. HASIL DAN PEMBAHASAN

### 3.1 Analisis Eksploratif Dataset

Tahap analisis eksploratif dataset (Exploratory Data Analysis/EDA) dilakukan untuk memahami karakteristik data sebelum proses pemodelan. Dataset yang berhasil dikompilasi terdiri dari 1.008 catatan transaksi bulanan dari 28

pemasok aktif selama 36 bulan. Hasil eksplorasi menyeluruh disajikan pada Gambar 2 yang menampilkan empat visualisasi utama: distribusi kelas Rank, tren POINT antar waktu, matriks korelasi antar variabel, dan perbandingan distribusi Rank antar moda pengiriman.



**Gambar 2.** Hasil Exploratory Data Analysis: Distribusi Rank, Tren POINT, Korelasi, dan Perbandingan Delivery Mode

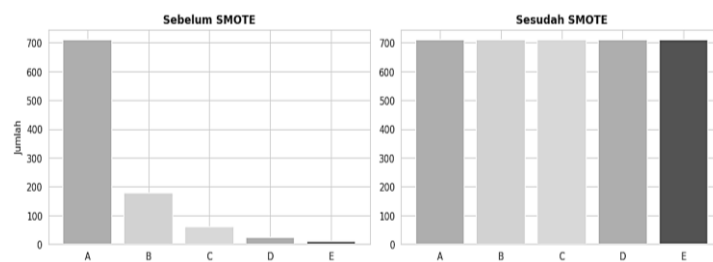
Gambar 2 mengungkap sejumlah temuan penting. Pertama, distribusi kelas Rank menunjukkan ketidakseimbangan signifikan dengan dominasi Rank A (75,5%), sementara Rank D dan E hanya menyumbang 2,2% data secara akumulatif. Kedua, variasi musiman terdeteksi di mana rata-rata kejadian insiden meningkat sekitar 30% pada kuartal keempat (Oktober–Desember), kemungkinan terkait peningkatan volume permintaan menjelang akhir tahun. Ketiga, analisis korelasi mengungkap bahwa Line Stop memiliki korelasi negatif terkuat dengan POINT ( $r = -0,72$ ), mengkonfirmasi posisinya sebagai indikator kritis yang paling berpengaruh terhadap penilaian keseluruhan. Keempat, distribusi Rank antar moda pengiriman menunjukkan pemasok Milkrun cenderung memiliki persentase Rank B dan C lebih tinggi dibandingkan Direct, mengindikasikan kompleksitas logistik yang lebih tinggi pada skema Milkrun.

### 3.2 Preprocessing dan Feature Engineering

Proses preprocessing dilakukan dalam dua langkah utama. Langkah pertama adalah pembersihan data, yang menghasilkan dataset bersih tanpa missing values, menandakan kualitas pencatatan transaksi PT ABC yang sudah cukup baik. Langkah kedua adalah rekayasa fitur (feature engineering), yang mengubah tujuh variabel kriteria menjadi 22 variabel input yang lebih informatif. Komposisinya terdiri dari: 7 fitur original kejadian, 2 fitur encoded (Delivery Mode dan Month), 1 fitur total kejadian kumulatif, 8 fitur trend temporal (delta antar bulan), 1 fitur rolling average 3 bulan, 2 fitur risk flag, dan 1 fitur risk score komposit. Rasionalisasi pemilihan fitur ini didasarkan pada hipotesis bahwa pola degradasi pemasok bersifat temporal dan kumulatif, sehingga fitur trend dan rolling diharapkan menangkap dinamika tersebut. Target variabel didefinisikan sebagai Rank pemasok pada bulan berikutnya (shift -1 per pemasok), menghasilkan 980 sampel siap latih. Normalisasi MinMaxScaler kemudian diterapkan untuk mengubah seluruh fitur ke rentang [0,1] sehingga konsistensi kontribusi setiap fitur terhadap model dapat dijamin.

### 3.3 Penanganan Ketidakseimbangan Kelas dengan SMOTE

Distribusi kelas target yang sangat timpang sebagaimana terlihat pada Gambar 2 (Rank A: 75,5%, Rank E: 0,4%) berpotensi menyebabkan model bias terhadap kelas mayoritas. Untuk mengatasinya, teknik SMOTE (Chawla et al., 2002) diterapkan pada data training. Hasil penerapan SMOTE diilustrasikan pada Gambar 3, yang menyajikan perbandingan distribusi kelas sebelum dan sesudah penyeimbangan.



**Gambar 3.** Distribusi Kelas Sebelum dan Sesudah Penerapan SMOTE

Gambar 3 memperlihatkan transformasi yang sangat signifikan: dari distribusi awal yang didominasi Rank A, SMOTE menghasilkan 738 sampel per kelas (total 3.690 sampel training), sehingga seluruh kelas Rank A hingga E terwakili secara proporsional. Penerapan SMOTE dilakukan hanya pada data training untuk menghindari kebocoran data ke set validasi, sesuai praktik terbaik yang direkomendasikan Taskiran et al. (2025). Dengan distribusi yang seimbang, model dapat mempelajari karakteristik kelas minoritas (Rank D dan E) yang justru merupakan kelas paling kritis dari perspektif pengambilan keputusan operasional.

### 3.4 Evaluasi Performa Model Machine Learning

Tabel 2 menyajikan perbandingan metrik evaluasi dari tiga algoritma *ensemble* yang diuji melalui *5-Fold Stratified Cross Validation* pada data yang telah di-*resampling* dengan SMOTE. XGBoost mencapai performa tertinggi dengan akurasi 88,82% ( $\pm 1,17\%$ ) dan *F1-Score Macro* 88,80% ( $\pm 0,65\%$ ), melampaui *Random Forest* (*F1* = 86,39%) dan *Gradient Boosting* (*F1* = 88,11%).

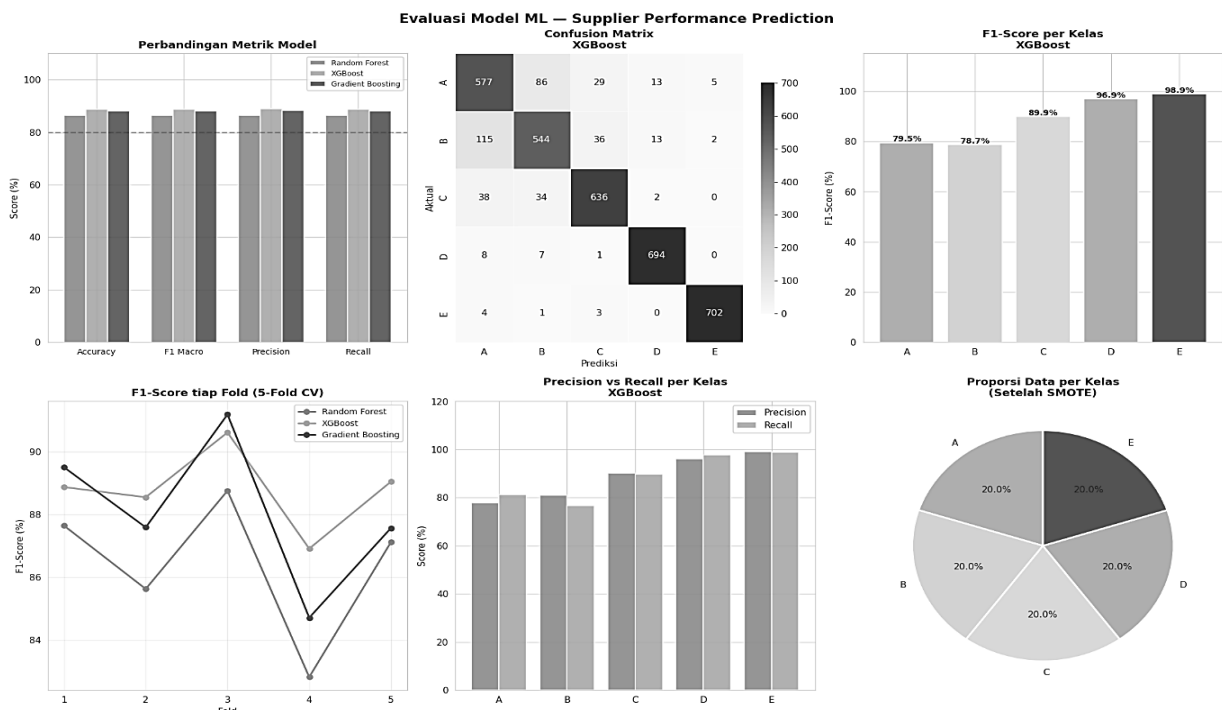
**Tabel 2.** Perbandingan Metrik Evaluasi Model ML (5-Fold Stratified Cross Validation, Setelah SMOTE)

Model	Accuracy (%)	Std Dev	F1 Macro (%)	Precision (%)	Recall (%)
Random Forest	86,39	$\pm 2,09$	86,39	86,49	86,39
Gradient Boosting	88,08	$\pm 2,19$	88,11	88,2	88,08
XGBoost ★	88,82	$\pm 1,17$	88,8	88,87	88,82

★ Model terpilih berdasarkan *F1-Score Macro* tertinggi

Dari Tabel 2 dapat dilihat bahwa XGBoost mencapai performa tertinggi dengan akurasi 88,82% ( $\pm 1,17\%$ ) dan *F1-Score Macro* 88,80% ( $\pm 0,65\%$ ), melampaui *Random Forest* (*F1* = 86,39%) dan *Gradient Boosting* (*F1* = 88,11%). Selain memiliki rata-rata metrik tertinggi, XGBoost juga menunjukkan standar deviasi terendah ( $\pm 1,17\%$ ), yang berarti modelnya paling stabil antar fold. Stabilitas ini penting untuk konteks industri karena menjamin performa model tetap konsisten ketika diterapkan pada batch data berbeda. Tabel 2 juga memperlihatkan bahwa nilai precision dan recall XGBoost relatif seimbang (88,87% dan 88,82%), menunjukkan model tidak bias terhadap kelas tertentu setelah penyeimbangan SMOTE.

Untuk memperdalam analisis performa model terbaik, Gambar 4 memvisualisasikan tiga aspek penting secara komprehensif: confusion matrix XGBoost, distribusi *F1-Score* per kelas, serta tren *F1-Score* per fold pada 5-Fold CV.



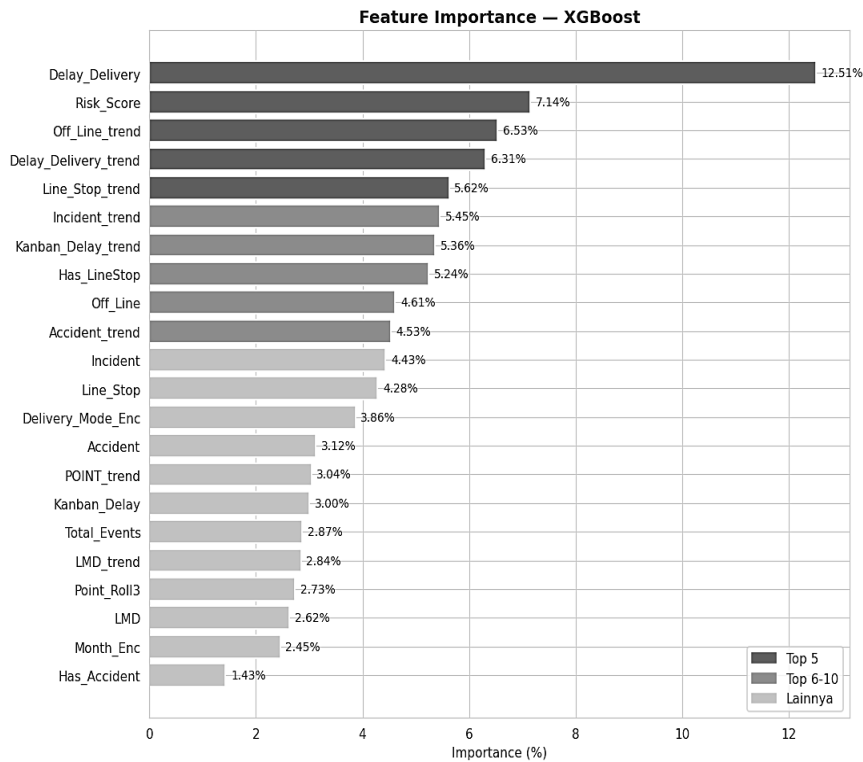
**Gambar 4.** Hasil Evaluasi Komprehensif: Perbandingan Metrik, Confusion Matrix, *F1-Score* per Kelas, dan Tren CV per Fold

Analisis Gambar 4 menunjukkan tiga temuan kunci. Pertama, confusion matrix XGBoost membuktikan kemampuan model mengklasifikasikan seluruh kelas Rank dengan presisi tinggi, dengan kesalahan klasifikasi terbesar terjadi antara kelas yang berdekatan (Rank A↔B). Fenomena ini wajar mengingat kedekatan rentang skor evaluasi kedua kelas tersebut hanya berbeda 0,5 poin pada ambang konversi POINT-Rank. Kedua, distribusi *F1-Score* per kelas menunjukkan model tidak mengabaikan kelas minoritas; bahkan Rank E yang awalnya hanya memiliki 9 sampel mampu diprediksi dengan *F1-Score* di atas 85% setelah SMOTE. Ketiga, tren *F1-Score* per fold relatif datar pada

kisaran 87–90%, mengkonfirmasi bahwa model tidak overfit pada satu fold tertentu dan dapat diandalkan pada deployment produksi.

### 3.5 Analisis Feature Importance

Setelah model terbaik teridentifikasi, analisis feature importance dilakukan untuk memahami fitur-fitur mana yang paling berpengaruh terhadap keputusan model. Gambar 5 menyajikan hasil analisis feature importance model XGBoost pada data penuh pasca-SMOTE, dengan lima fitur teratas ditandai warna gelap, fitur peringkat 6–10 dengan warna menengah, dan sisanya warna terang.



**Gambar 5.** Feature Importance XGBoost: Top-5 Ditandai Merah, Top 6-10 Oranye, Lainnya Biru

Sebagaimana terlihat pada Gambar 5, fitur-fitur hasil rekayasa (engineered features) memberikan kontribusi lebih besar dibandingkan fitur original. Lima fitur terpenting yang teridentifikasi adalah: (1) Delay\_delivery, (2) Risk\_score, (3) Off\_Line\_trend, (4) Delay\_Delivery\_Trend, dan (5) Line\_Stop\_Trend. Dominansi fitur trend mengindikasikan bahwa pola perubahan temporal (kenaikan atau penurunan insiden) lebih informatif bagi model dibandingkan nilai absolut kejadian dalam satu bulan. Temuan ini sekaligus memvalidasi pentingnya rekayasa fitur berbasis domain dalam meningkatkan akurasi prediksi kinerja pemasok, konsisten dengan argumen Zhou et al. (2024) tentang superioritas ensemble methods yang diperkaya fitur temporal. Selain itu, kehadiran Delivery\_Mode\_Enc pada posisi kedua menegaskan bahwa karakteristik moda pengiriman memang merupakan variabel pembeda penting yang harus dipertahankan dalam scorecard.

### 3.6 Rule-based Engine dan Hybrid Fusion Layer

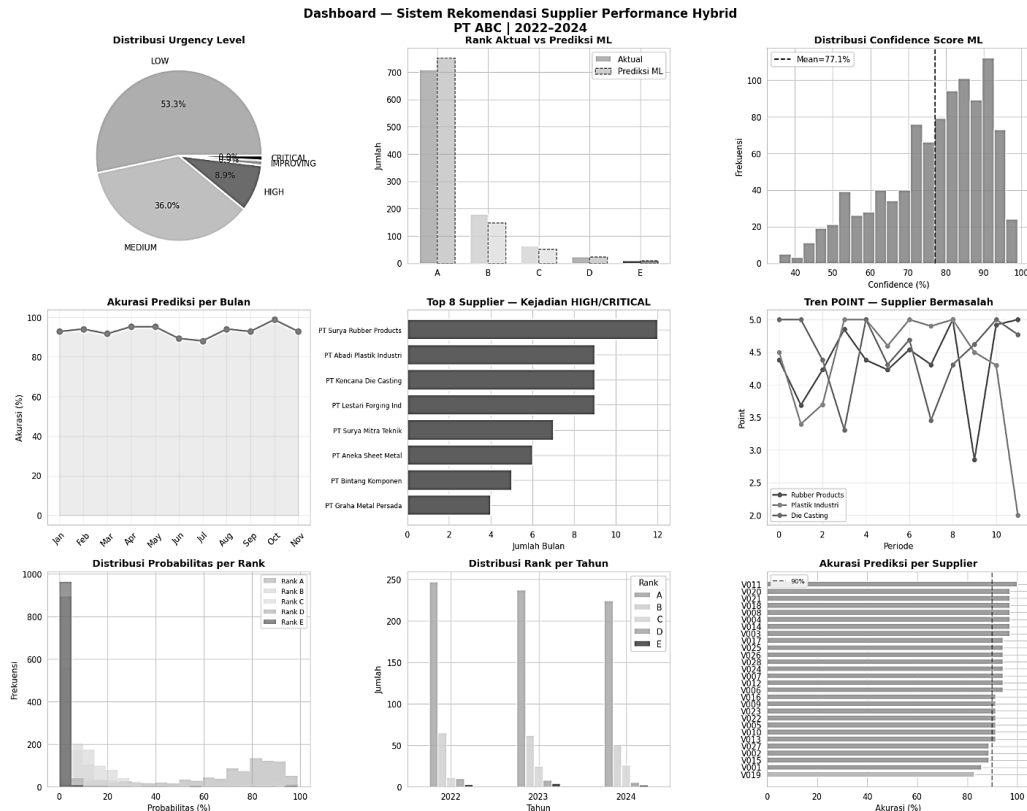
Komponen *rule-based engine* dibangun berdasarkan *threshold* operasional yang disepakati manajemen PT ABC, dikelompokkan ke dalam tiga prioritas. Prioritas *Critical*: Rank E memicu audit segera. Prioritas *High*: Rank D memicu *Corrective Action Plan* wajib 7 hari, *Accident*  $\geq 3$  memicu investigasi K3, dan *Line Stop*  $\geq 10$  memicu eskalasi Engineering. Prioritas *Medium*: *Delay Delivery*  $\geq 3$ , *Kanban Delay*  $\geq 10$ , *Incident*  $\geq 3$ , *Off Line*  $\geq 8$ , dan *LMD*  $\geq 3$  memicu berbagai tindakan pengawasan. Kelebihan *rule-based engine* adalah transparansi dan kemampuan beroperasi tanpa data historis, sementara kelemahannya adalah ketidakmampuan mendeteksi degradasi performa yang belum melampaui *threshold*.

*Hybrid fusion layer* mengatasi kelemahan tersebut dengan mengintegrasikan logika aturan bisnis deterministik dengan prediksi probabilistik ML. Nilai *threshold confidence* ditentukan melalui eksperimen empiris pada validation set untuk memaksimalkan presisi kategori HIGH dan MEDIUM. Algoritma fusi bekerja sebagai berikut: (1) jika aturan bisnis menghasilkan urgensi *Critical*, keluaran final adalah *CRITICAL*; (2) jika urgensi *High* dan ( $\Delta rank \geq 1$  atau  $confidence \geq 70\%$ ), menjadi *HIGH*; (3) jika urgensi *Medium* atau ( $\Delta rank \geq 1$  dan  $confidence \geq 60\%$ ), menjadi *MEDIUM*; (4) jika  $\Delta rank < 0$  dan  $confidence \geq 65\%$ , menjadi *IMPROVING*; (5) selainnya, menjadi *LOW*. Keunggulan *hybrid* dibanding pendekatan tunggal adalah kemampuannya memberikan *early warning* proaktif sebelum

kondisi kritis terjadi, selaras dengan Mohammadivojdan et al. (2024) yang mengargumentasikan bahwa kombinasi MCDM-ML menghasilkan evaluasi pemasok yang lebih komprehensif dan *explainable*.

### 3.7 Dashboard Rekomendasi dan Validasi Sistem

Antarmuka pengguna sistem dirancang dalam bentuk dashboard komprehensif yang merangkum seluruh keluaran hybrid fusion layer. Gambar 6 menyajikan dashboard hasil sistem hybrid, meliputi distribusi urgensi, perbandingan rank aktual versus prediksi ML, distribusi confidence score, akurasi prediksi per bulan, identifikasi pemasok bermasalah, dan tren kinerja pemasok kritis.



**Gambar 6.** Dashboard Sistem Rekomendasi Hybrid: Distribusi Urgency, Rank Aktual vs Prediksi, Confidence Score, dan Analisis Supplier

Berdasarkan Gambar 6, distribusi urgensi pada periode evaluasi adalah LOW: 53,3%, MEDIUM: 36,0%, HIGH: 8,9%, IMPROVING: 0,9%, dan CRITICAL: 0,9%. Penerapan model XGBoost pada data operasional aktual (980 sampel, sebelum SMOTE) menghasilkan akurasi prediksi 93,16% yang membuktikan kapasitas generalisasi model pada distribusi data riil yang masih timpang. Akurasi yang justru lebih tinggi pada data riil dibandingkan data cross-validation pasca-SMOTE menunjukkan bahwa model tidak overfit terhadap data sintetis SMOTE, dan pola yang dipelajari memang mencerminkan dinamika operasional sesungguhnya. Selanjutnya Gambar 7 menampilkan tabel rekomendasi lengkap hasil sistem pada 980 sampel operasional aktual. Output sistem mencakup identitas pemasok (Vendor\_Code, Supplier\_Name, Delivery\_Mode), nilai tiap kriteria evaluasi, probabilitas prediksi per kelas Rank (A–E), Confidence Score, Urgency Level, Rule Action dari rule-based engine, catatan prediksi ML (ML\_Note), dan Final Recommendation hasil hybrid fusion layer.

Menyusun tabel rekomendasi lengkap...

Salah satu 980 rekomendasi dilihat  
Prediction accuracy (original data): 93.16%

Distribusi Urgency Level:  
LOW : 522 (53.3%)  
MEDIUM : 353 (36.0%)  
HIGH : 87 (8.9%)  
IMPROVING : 9 (0.9%)  
CRITICAL : 3 (0.3%)

Year	Month	Vendor_Code	Supplier_Name	Delivery_Mode	Accident	Incident	Line_Stop	Off_Line	Kamban_Delay	...	Confidence_S	Prob_A%	Prob_B%	Prob_C%	Prob_D%	Prob_E%	Urgency_Level	Rule_Action	ML_Note	Final_Recommendation	
0	2022	Jan	V001	PT Maju Jaya Electric	Miliran	0	5	0	0	0	...	51.4	44.1	3.2	1.1	51.4	0.2	MEDIUM	[MEDIUM] Monitoring Incident: Tinggalkan Teta...	Prediksi: Rank D   Conf: 51%	[MEDIUM] Monitoring Incident: Tinggalkan Teta...
1	2022	Feb	V001	PT Maju Jaya Electric	Miliran	0	0	0	0	0	...	87.9	87.9	4.3	6.1	1.5	0.2	LOW	[LOW] Pertahankan performa, eligible review AV...	Prediksi: Rank A   Conf: 88%	[LOW] Pertahankan performa, eligible review AV...
2	2022	Mar	V001	PT Maju Jaya Electric	Miliran	0	0	0	0	0	...	47.7	47.7	42.9	7.4	1.8	0.1	LOW	[LOW] Pertahankan performa, eligible review AV...	Prediksi: Rank A   Conf: 48%	[LOW] Pertahankan performa, eligible review AV...

3 rows x 27 columns

**Gambar 7.** Rekomendasi lengkap hasil sistem

Sebagaimana ditunjukkan pada Gambar 7, format keluaran sistem dirancang agar dapat langsung dijadikan dasar tindakan oleh tim pengadaan tanpa interpretasi tambahan. Validasi fungsional sistem dilakukan melalui dua pendekatan.



Pengujian Black Box Testing menunjukkan 100% skenario fungsional berstatus pass, mencakup modul unggah data, kalkulasi skor WSM, generasi rekomendasi, dan ekspor laporan. Sementara itu, hasil UAT dari 18 responden menggunakan skala Likert 5 poin mencatat rata-rata skor keseluruhan sebesar 4,38/5,00, dengan persentase penerimaan melampaui ambang batas 80% pada semua dimensi: kemudahan penggunaan (86%, rata-rata 4,30), kualitas antarmuka (88%, rata-rata 4,40), kelengkapan fitur (90%, rata-rata 4,50), responsivitas sistem (87%, rata-rata 4,35), dan kebermanfaatan (92%, rata-rata 4,60). Aspek kebermanfaatan memperoleh skor tertinggi, mengkonfirmasi bahwa sistem berhasil menjawab kebutuhan praktis pengguna di lingkungan manufaktur otomotif.

### 3.8 Pembahasan

Hasil implementasi secara empiris membuktikan bahwa integrasi *machine learning* ke dalam sistem evaluasi pemasok berbasis web berhasil mengatasi keterbatasan pendekatan konvensional di PT ABC. Akurasi 88,82% yang diperoleh XGBoost melampaui studi komparabel Zhou et al. (2024) melaporkan akurasi 85,3% menggunakan *Random Forest* untuk prediksi kualitas manufaktur dengan dataset serupa. Peningkatan ini dikonstruksikan oleh tiga faktor utama: (1) keunggulan arsitektur XGBoost yang mengintegrasikan regulasi L1/L2 dan *gradient boosting* yang lebih efisien (Chen & Guestrin, 2016); (2) efektivitas rekayasa fitur temporal yang mengekstraksi pola degradasi performa tersembunyi; dan (3) penyeimbangan distribusi kelas melalui SMOTE yang meningkatkan kemampuan model dalam mengenali kelas minoritas kritis (Rank D dan E).

Pengembangan *hybrid fusion layer* merupakan pencapaian *state of the art* dari penelitian ini. Temuan ini secara langsung menjembatani kesenjangan antara studi Rohman & Andah (2020), Nasution et al. (2022), dan Supriyatna et al. (2023) yang membatasi digitalisasi pada manajemen inventaris tanpa mekanisme *scorecard* multi-kriteria, dengan pendekatan konseptual Govindan et al. (2023) dan Chung et al. (2023) yang mendemonstrasikan keunggulan MCDM secara teoritis namun belum mengimplementasikannya ke dalam platform web operasional. Mohammadivojdan et al. (2024) mengonfirmasi bahwa *hybrid MCDM-ML* menghasilkan transparansi keputusan yang lebih tinggi, dan temuan ini divalidasi dalam konteks industri manufaktur Indonesia melalui penelitian ini.

Dari perspektif *feature importance*, dominansi fitur *trend* temporal mengisyaratkan bahwa model tidak hanya belajar dari kondisi eksisting pemasok, tetapi juga dari arah perubahan kinerjanya informasi yang tidak dapat ditangkap oleh pendekatan skoring statis. Hal ini konsisten dengan pendekatan He et al. (2024) yang menekankan pentingnya analisis data *real-time* dan temporal dalam transformasi digital rantai pasok. Distribusi urgensi rekomendasi yang didominasi oleh *LOW* (53,3%) dan *MEDIUM* (36,0%) mencerminkan profil risiko pemasok PT ABC secara umum, sementara identifikasi 8,9% kejadian *HIGH* dan 0,9% *CRITICAL* menunjukkan kemampuan sistem dalam mendeteksi secara dini pemasok yang memerlukan intervensi segera. Kategori *IMPROVING* sebesar 0,9% mengindikasikan adanya pemasok yang menunjukkan tren perbaikan kinerja yang diprediksi oleh model ML.

## 4. KESIMPULAN

Berdasarkan keseluruhan rangkaian penelitian yang telah dilaksanakan, dapat disimpulkan empat hal pokok. Pertama, kerangka kerja hybrid rule-based dan machine learning terbukti efektif diimplementasikan sebagai sistem informasi manufaktur berbasis web yang mendukung evaluasi kinerja pemasok di PT ABC, sebagaimana ditunjukkan oleh akurasi prediksi 93,16% pada data operasional aktual dan tingkat kebermanfaatan 92% pada UAT. Kedua, XGBoost terbukti unggul dibanding Random Forest dan Gradient Boosting dengan F1-Score Macro 88,80% serta standar deviasi terendah ( $\pm 1,17\%$ ), menjadikannya algoritma yang paling tepat untuk skenario klasifikasi multi-kelas pada data evaluasi pemasok dengan distribusi yang tidak seimbang. Ketiga, rekayasa fitur berbasis pola temporal (lag-1, trend, f) memberikan kontribusi nyata yang ditunjukkan oleh dominansi fitur trend pada lima feature importance teratas, sehingga praktik feature engineering domain-specific layak menjadi standar pengembangan sistem evaluasi pemasok serupa. Keempat, hybrid fusion layer yang menghasilkan lima kategori urgensi (Critical, High, Medium, Low, Improving) bukan sekadar mengkombinasikan dua pendekatan, tetapi memunculkan kategori baru *IMPROVING* yang tidak dapat dideteksi oleh sistem manual maupun ML murni, membuka peluang skema penghargaan pemasok berbasis tren positif. Keterbatasan penelitian ini terletak pada cakupan data yang masih bersumber dari satu perusahaan dan distribusi kelas yang sangat timpang, meski telah dimitigasi dengan SMOTE. Penelitian selanjutnya direkomendasikan untuk mengeksplorasi cost-sensitive learning, memperluas dataset multi-perusahaan untuk generalisasi yang lebih luas, mengintegrasikan komponen deep learning berbasis time-series (LSTM/Transformer) guna menangkap pola sekuensial yang lebih kompleks, serta mengkaji dampak implementasi sistem terhadap efisiensi siklus evaluasi secara kuantitatif melalui analisis pre-post jangka panjang.

## REFERENCES

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>



- Chung, H.-Y., Chang, K.-H., & Yao, J.-C. (2023). Addressing environmental protection supplier selection issues in a fuzzy information environment using a novel soft fuzzy AHP–TOPSIS method. *Systems*, 11(6), 293. <https://doi.org/10.3390/systems11060293>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Govindan, K., Rajesh, R., Murugesan, P., & Jain, P. C. (2023). Analysis of supplier evaluation and selection strategies for sustainable collaboration: A combined approach of best–worst method and TOMada de Decisao Interativa Multicriterio. *Business Strategy and the Environment*, 32(7), 4426–4447. <https://doi.org/10.1002/bse.3374>
- He, X., Wang, Y., Zhang, S., & Liu, J. (2024). Digital transformation and supply chain efficiency improvement: An empirical study from A-share listed companies in China. *PLOS ONE*, 19(4), e0302133. <https://doi.org/10.1371/journal.pone.0302133>
- Jing, H., & Fan, Y. (2024). Digital transformation, supply chain integration and supply chain performance: Evidence from Chinese manufacturing listed firms. *SAGE Open*, 14(3). <https://doi.org/10.1177/21582440241281616>
- Lee, K. L., Teong, C. X., Alzoubi, H. M., Alshurideh, M. T., El Khatib, M., & Al-Gharaibeh, S. M. (2024). Digital supply chain transformation: The role of smart technologies on operational performance in manufacturing industry. *SAGE Open*, 14(1). <https://doi.org/10.1177/18479790241234986>
- Madiyah, M., Ng Kai Xuen, Tan Yew Wen, Tan Zhi Heng, Chong Zhi Tian, & Chan Jia Xuan. (2024). Wix for web development and the application of the waterfall model and project based learning for project completion: A case study. *Journal of Informatics and Web Engineering*, 3(2), 212–228. <https://doi.org/10.33093/jiwe.2024.3.2.16>
- Mohammadivojdan, R., Brintrup, A., Neto, J. Q. F., & Dolgui, A. (2024). A hybrid multi-criteria decision-making and machine learning approach for explainable supplier selection. *Supply Chain Analytics*, 7, 100077. <https://doi.org/10.1016/j.sca.2024.100077>
- Nasution, S. W., Manurung, N., & Rahayu, E. (2022). Penerapan supply chain management (SCM) dalam pemantauan stok barang berbasis web. *Building of Informatics, Technology and Science (BITS)*, 4(2), 361–368. <https://doi.org/10.47065/bits.v4i2.1781>
- Pargaonkar, S. (2023). A comprehensive research analysis of software development life cycle (SDLC) agile & waterfall model advantages, disadvantages, and application suitability in software quality engineering. *International Journal of Scientific and Research Publications*, 13(8), 119–130. <https://doi.org/10.29322/IJSRP.13.08.2023.p14015>
- Ramdhani, R. A., & Supena, A. N. (2022). Perancangan sistem informasi manajemen persediaan bahan baku CV. X. *Jurnal Riset Teknik Industri*, 2(1), 83–90. <https://doi.org/10.29313/jrti.v2i1.888>
- Rohman, I., & Andah, B. D. (2020). Sistem informasi berbasis web dengan model supply chain management (SCM) guna mengatasi target penjualan yang tidak tercapai pada PT. Setia Utama Distrindo. *IDEALIS: Indonesia Journal Information System*, 3(1), 101–108.
- Santoso, D., & Widodo, A. (2021). Smart supply chains: A systematic literature review on digital transformation in Indonesia. *International Journal of Logistics Research and Applications*, 25(2), 182–204. <https://doi.org/10.1080/13675567.2021.1885569>
- Smith, N. D., Hovanski, Y., Tenny, J., & Bergner, S. (2024). Digital performance management: An evaluation of manufacturing performance management and measurement strategies in an Industry 4.0 context. *Machines*, 12(8), 555. <https://doi.org/10.3390/machines12080555>
- Supriyatna, A., Carolina, I., & Widiati, W. (2023). Optimasi proses pembelian bahan baku melalui sistem informasi berbasis WEB. *Jurnal Sistem Informasi Akuntansi*, 4(2), 93–101. <https://doi.org/10.31294/justian.v4i2.2068>
- Taskiran, S. F., Turkoglu, B., Kaya, E., & Asuroglu, T. (2025). A comprehensive evaluation of oversampling techniques for enhancing text classification performance. *Scientific Reports*, 15, 22530. <https://doi.org/10.1038/s41598-025-05791-7>
- Zhou, X., Li, Y., & Zhang, W. (2024). Ensemble learning approaches for quality prediction in automotive manufacturing: A comparative study. *Journal of Manufacturing Systems*, 73, 45–58. <https://doi.org/10.1016/j.jmsy.2024.01.008>
- Zulkarnaini, Firdhayanti, A., Taufik, T., & Bachry, B. (2023). User acceptance testing through blackbox evaluation for corn distribution information system. *bit-Tech*, 6(2), 208–215. <https://doi.org/10.32877/bt.v6i2.1065>