



Integrasi Feature Engineering dan SMOTE pada Algoritma Random Forest untuk Prediksi Kerusakan Chip RFID di Industri Sel Surya

Haeruddin*, Franklin Winata, Muhammad Ilham Ashiddiq Tresnawan, Gautama Wijaya, Heru Wijayanto Aripardono

Fakultas Ilmu Komputer, Program Studi Teknologi Informasi, Universitas Internasional Batam, Batam Baloi-Sei Ladi, Jl. Gajah Mada, Tiban Indah, Kec. Sekupang, Kota Batam, Kepulauan Riau, Indonesia
Email: ¹*haeruddin@uib.ac.id, ²2232063.franklin@uib.edu ³ilham.tresnawan@uib.ac.id, ⁴gautama.wijaya@uib.ac.id, ⁵heru.wijayanto@uib.ac.id

Email Penulis Korespondensi: haeruddin@uib.ac.id

Submitted: 26/12/2025; Accepted: 31/01/2026; Published: 31/01/2026

Abstrak—Industri manufaktur sel surya menuntut proses produksi yang cepat, presisi, dan didukung oleh integritas data yang tinggi. Kerusakan chip RFID pada flower basket dapat menyebabkan kegagalan pembacaan data dan mengganggu efisiensi serta kelancaran proses produksi. Penelitian ini bertujuan untuk memprediksi status chip, yaitu dalam kondisi aman atau rusak, berdasarkan berbagai parameter proses, meliputi suhu perendaman, kelembapan ruang, tekanan proses, getaran mesin, kecepatan pengering, waktu pemanasan dan pendinginan, temperatur pembakaran, jumlah pemakaian, serta kondisi pembacaan RFID. Pendekatan feature engineering diterapkan untuk membangun fitur turunan yang lebih representatif, sementara SMOTE digunakan untuk mengatasi ketidakseimbangan kelas data. Penelitian ini berfokus pada pengembangan model prediksi menggunakan metode Random Forest untuk mengidentifikasi variabel proses yang paling berpengaruh terhadap potensi kerusakan chip. Data yang digunakan berasal dari catatan historis proses produksi pada pabrik sel surya. Hasil penelitian menunjukkan bahwa kombinasi beberapa parameter proses memiliki kontribusi signifikan terhadap potensi kerusakan chip, serta model Random Forest mampu menghasilkan kinerja prediksi yang baik dalam mengklasifikasikan kondisi chip. Temuan ini berpotensi dimanfaatkan sebagai sistem peringatan dini untuk mendeteksi risiko kerusakan chip sebelum berdampak pada proses produksi. Dengan implementasi yang tepat, model prediksi ini diharapkan dapat mendukung tindakan pencegahan, meningkatkan integritas data, serta meminimalkan gangguan dalam alur manufaktur sel surya..

Kata Kunci: Manufaktur Sel Surya; Prediksi Kerusakan; RFID; Feature Engineering; Random Forest; SMOTE

Abstract—The electronics industry, particularly solar cell manufacturing, demands production processes that are fast, precise, and supported by high data integrity. One critical component in the production flow is the chip embedded in the flower basket, which functions to store and transmit data through an RFID system. Damage to the chip can lead to information loss, tag reading failures, and disruptions in production efficiency and continuity. This study aims to predict chip status, classified as either normal or damaged, based on various process parameters, including immersion temperature, ambient humidity, process pressure, machine vibration, drying speed, heating and cooling duration, firing temperature, usage frequency, and RFID reading conditions. A feature engineering approach is applied to construct more representative derived features, while SMOTE is utilized to address class imbalance in the dataset. This study focuses on developing a predictive model using the Random Forest method to identify the most influential process variables related to chip damage risk. The data used in this study are obtained from historical production process records of a solar cell manufacturing plant. The results indicate that combinations of multiple process parameters significantly contribute to the potential risk of chip damage, and the Random Forest model demonstrates good predictive performance in classifying chip conditions. These findings suggest that the proposed model can serve as an early warning system to detect chip damage risks before they impact production processes. With proper implementation, the predictive model is expected to support preventive actions, enhance data integrity, and minimize disruptions in the solar cell manufacturing workflow.

Keywords: Solar Cell Manufacturing; Damage Prediction; RFID; Feature Engineering; Random Forest; SMOTE

1. PENDAHULUAN

Industri pembuatan sel surya terus berkembang seiring meningkatnya kebutuhan akan energi terbarukan, sehingga sistem pemantauan material yang akurat dan andal menjadi semakin penting. Salah satu teknologi yang banyak diterapkan untuk tujuan tersebut adalah Radio Frequency Identification (RFID)[1]. Teknologi RFID memanfaatkan komunikasi gelombang radio untuk memungkinkan pertukaran data antara RFID reader dan RFID tag, tanpa memerlukan kontak fisik atau garis pandang langsung. Ketika reader memancarkan sinyal radio, tag akan merespon dengan mengirimkan informasi identitas yang tersimpan pada chip, sehingga memungkinkan proses identifikasi dan pelacakan material secara otomatis dan real-time di berbagai tahapan produksi.[2]

Dalam konteks manufaktur sel surya, chip RFID dipasangkan pada keranjang bunga (flower basket) yang digunakan untuk membawa wafer melalui berbagai proses produksi, mulai dari tahap awal hingga proses akhir. Setiap kali flower basket melewati titik pembacaan (reader), data identitas dan status proses akan direkam oleh sistem manufaktur. Oleh karena itu, keandalan chip RFID sangat berpengaruh terhadap kelancaran alur informasi, akurasi pelacakan material, serta integritas data produksi secara keseluruhan[3], [4], yang menjadi dasar utama pengambilan keputusan operasional di lingkungan manufaktur.

Kerusakan chip RFID dapat menyebabkan gangguan pembacaan data, kehilangan informasi proses, serta kegagalan pelacakan material secara otomatis. Pada sistem produksi bervolume tinggi, kondisi ini berpotensi meningkatkan waktu henti produksi, menurunkan efisiensi lini manufaktur, dan memicu kebutuhan verifikasi



manual yang menambah biaya operasional. Selain itu, hilangnya data proses akibat kerusakan chip RFID juga dapat menurunkan tingkat traceability produk dan menyulitkan proses analisis kualitas serta penelusuran cacat produksi.

Penelitian terdahulu menunjukkan bahwa performa RFID sangat dipengaruhi oleh kondisi operasional di lingkungan industri. Studi oleh [5] mengkaji pengaruh suhu tinggi dan beban stres termal terhadap kinerja antena RFID, dan menemukan bahwa paparan suhu ekstrem dapat mempercepat degradasi material antena serta menurunkan sensitivitas pembacaan secara signifikan. Penelitian lain oleh [6] mengevaluasi dampak tekanan mekanis dan getaran mesin terhadap RFID tag pasif, yang menunjukkan bahwa kondisi mekanis yang fluktuatif dapat menyebabkan penurunan performa pembacaan dan perubahan karakteristik antena secara bertahap. Sementara itu, penelitian [7] menganalisis pengaruh lingkungan industri yang keras terhadap stabilitas sinyal dan keandalan komunikasi RFID, dengan fokus pada karakteristik fisik tag, interaksi dengan material logam, perubahan suhu, dan interferensi lingkungan yang menurunkan kualitas pembacaan. Penelitian [8] menunjukkan bahwa penerapan algoritma machine learning pada sistem sensor berbasis RFID mampu mengatasi permasalahan data yang tidak lengkap dan bersifat bising, sehingga meningkatkan akurasi estimasi dan keandalan data pada lingkungan operasional yang kompleks. Temuan ini mengindikasikan bahwa pendekatan machine learning berpotensi digunakan untuk meningkatkan reliabilitas sistem RFID pada aplikasi industri.

Gangguan pembacaan RFID dalam sistem manufaktur otomatis juga sering terjadi akibat adanya bahan penghalang, posisi tag yang berubah, atau interferensi dari peralatan produksi. Ketidakstabilan pembacaan RFID seperti ini dapat berakibat pada hilangnya data proses maupun kegagalan sistem pelacakan material, sehingga memerlukan pendekatan prediksi yang mampu mengidentifikasi potensi anomali sejak dini.[9], [10]

Seiring berkembangnya teknologi industri, machine learning telah menjadi pendekatan yang efektif dalam memprediksi kegagalan komponen elektronik dan memantau kualitas proses secara otomatis. Berbagai penelitian telah membuktikan bahwa model Random Forest dapat memberikan performa klasifikasi yang tinggi untuk analisis kualitas produksi, deteksi anomali, serta prediksi kerusakan komponen berbasis data sensor industri[11], [12][13]. Dengan kemampuan mengenali pola non-linear dan hubungan kompleks antar-parameter proses, Random Forest termasuk algoritma yang paling banyak digunakan pada industri cerdas (smart manufacturing).[14]

Namun, penelitian yang secara khusus memprediksi kerusakan chip RFID menggunakan berbagai fitur proses masih sangat terbatas. Sebagian besar studi terdahulu berfokus pada desain antena RFID atau dampak lingkungan secara umum, bukan pada pemodelan prediktif berbasis data proses. Oleh karena itu, penelitian ini bertujuan membangun model prediksi kerusakan chip RFID dengan memanfaatkan fitur proses produksi seperti suhu rendam, tekanan mesin, getaran peralatan, suhu pembakaran, kecepatan pengering, waktu pemanasan dan pendinginan, temperatur pembakaran, jumlah pemakaian serta kondisi pembacaan RFID sebagai indikator awal potensi kerusakan.

Untuk meningkatkan representasi informasi dari data proses, penelitian ini menerapkan teknik feature engineering, yaitu pembentukan fitur turunan yang dihasilkan dari kombinasi, rasio, maupun perubahan nilai antar-parameter proses[15]. Fitur turunan ini dirancang untuk menangkap pola tersembunyi dan hubungan non-linear yang tidak dapat direpresentasikan secara optimal oleh fitur asli, sehingga mampu meningkatkan akurasi prediksi pada aplikasi machine learning industri[16]. Selain itu, permasalahan ketidakseimbangan data ditangani menggunakan metode SMOTE (Synthetic Minority Oversampling Technique) dan variannya, yang terbukti efektif dalam meningkatkan performa model klasifikasi pada kasus kegagalan komponen dengan distribusi kelas tidak seimbang[17]. Penelitian modern juga menunjukkan bahwa penggabungan parameter proses, teknik penyeimbangan data, dan model Random Forest dapat meningkatkan kemampuan prediksi dalam aplikasi perawatan prediktif dan pemantauan kondisi industri[18].

Dengan mengintegrasikan data fitur proses, fitur turunan, dan algoritma machine learning modern, penelitian ini diharapkan mampu menghasilkan model prediksi kerusakan chip RFID yang akurat dan dapat digunakan sebagai sistem pendukung keputusan untuk meningkatkan efisiensi serta keandalan proses produksi pada industri sel surya.

2. METODOLOGI PENELITIAN

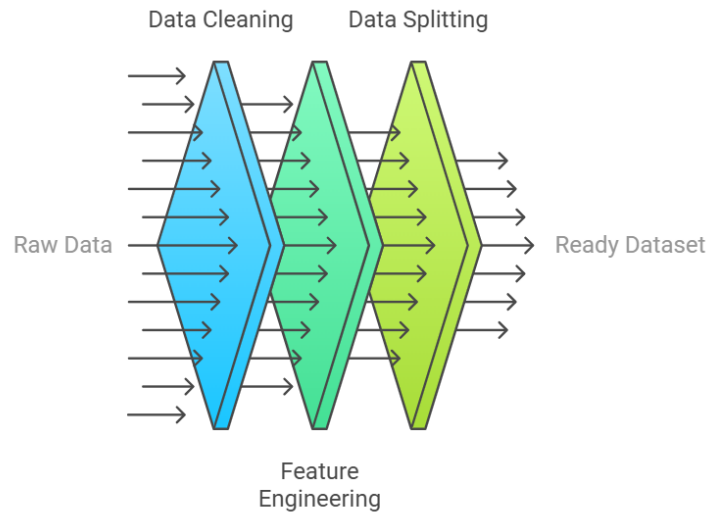
2.1 Desain Penelitian

Penelitian ini menggunakan metode kuantitatif berbasis machine learning untuk memprediksi risiko kerusakan chip RFID berdasarkan parameter proses produksi. Pendekatan ini dipilih karena machine learning mampu mempelajari pola kompleks serta hubungan non-linear antar variabel yang sulit diidentifikasi secara manual[19]. Algoritma Random Forest diterapkan sebagai model utama karena memiliki stabilitas tinggi, mampu menangani data berdimensi besar, dan telah terbukti efektif pada aplikasi pemantauan kualitas serta prediksi kerusakan di industri manufaktur modern[20], [21].

2.2 Tahapan Penelitian

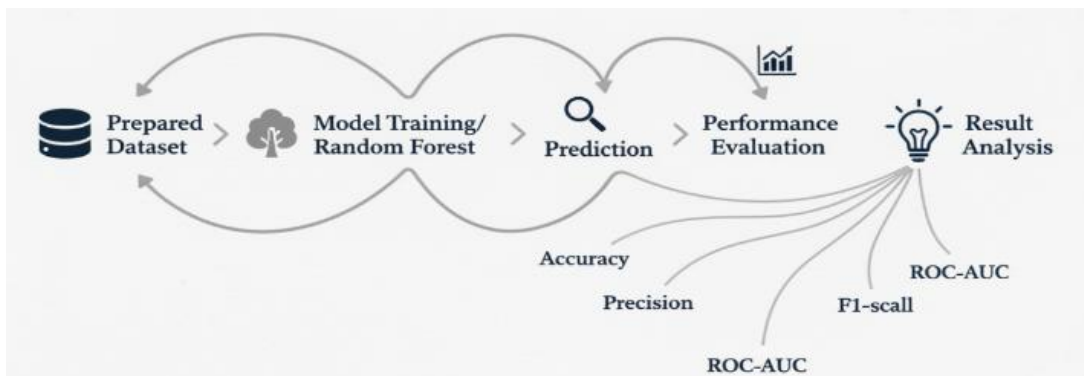
Penelitian ini dilakukan melalui beberapa tahapan yang tersusun secara sistematis dan dapat dikelompokkan ke dalam dua bagian utama, yaitu tahap pengolahan data (data processing) dan tahap pemodelan serta evaluasi

(modeling and evaluation). Alur tahapan pengolahan data ditunjukkan pada Gambar 1, sedangkan tahap pemodelan dan evaluasi dijelaskan secara terpisah pada bagian selanjutnya.



Gambar 1. Alur tahapan data processing pada penelitian.

Tahap pemodelan dan evaluasi dilakukan menggunakan ready dataset yang dihasilkan dari tahap pengolahan data. Pada tahap ini, model klasifikasi dibangun menggunakan algoritma Random Forest untuk memprediksi kerusakan chip RFID. Kinerja model kemudian dievaluasi menggunakan beberapa metrik, yaitu akurasi, precision, recall, F1-score, dan ROC-AUC, guna menilai kemampuan model dalam mendeteksi chip RFID yang mengalami kerusakan. Hasil evaluasi selanjutnya dianalisis untuk menilai efektivitas integrasi feature engineering dan SMOTE, yang kemudian digunakan sebagai dasar penarikan kesimpulan penelitian.



Gambar 2. Alur tahapan modeling and evaluation pada penelitian.

2.3 Data

Dataset yang digunakan terdiri dari dua jenis:

1. Data mentah yang berasal dari PT Nusa Solar Indonesia, mencakup:

Tabel 1. Data Fitur

No	Fitur	No	Fitur
1	Suhu Rendam	6	Waktu pemanasan
2	Kelembapan ruang	7	Waktu pendinginan
3	Tekanan proses	8	Temperatur pembakaran
4	Getaran mesin	9	Jumlah pemakaian
5	Kecepatan pengering	10	Kondisi pembacaan RFID

Fitur-fitur pada Tabel 1 mencakup parameter suhu, kelembapan, tekanan, getaran, serta waktu proses yang berpotensi memengaruhi kualitas dan keandalan chip RFID. Selain itu, variabel kondisi pembacaan RFID digunakan sebagai indikator awal untuk mengidentifikasi performa pembacaan chip dalam lingkungan industri[22].

2. Data yang telah diproses dengan feature engineering untuk meningkatkan kemampuan model dalam mengenali pola. Data disusun berdasarkan kombinasi matematis dari variabel mentah sehingga menghasilkan indikator

risiko yang lebih jelas. Parameter ini merupakan faktor yang berpotensi memengaruhi kinerja dan keandalan chip RFID, sebagaimana dijelaskan dalam studi mengenai dampak lingkungan industri terhadap performa tag RFID.[23], [24]

2.4 Feature Engineering

Tahap feature engineering dilakukan untuk mengekstraksi pola kritis dari data proses. Feature engineering dirancang berdasarkan hubungan fisik antar variabel dan potensi mekanisme kerusakan chip RFID. Pendekatan ini didukung oleh penelitian terbaru yang menunjukkan bahwa teknik feature engineering dapat meningkatkan kinerja model prediksi secara signifikan pada aplikasi industri[25]Seluruh fitur yang digunakan merupakan parameter proses yang tersedia sebelum tahap inspeksi akhir.

Fitur Delta_Suhu didefinisikan sebagai selisih antara suhu pembakaran dan suhu rendam, yang dirumuskan sebagai $F_1 = P - R$. Beban panas-lembap direpresentasikan oleh HumidityHeat dengan rumus $F_2 = R \times K$, sedangkan tekanan mekanis mesin dinyatakan melalui MachineStress dengan rumus $F_3 = G \times X \times V$. Laju pendinginan dinyatakan oleh CoolingRate dengan rumus $F_4 = \frac{H}{C}$ sebagai indikator perubahan suhu mendadak.

Kondisi rendaman ekstrem direpresentasikan oleh Suhu_Rendam_Too_High, yang bernilai 1 jika suhu rendam melebihi 90°C. Getaran berbahaya ditunjukkan oleh Getaran_Ekstrim, yang bernilai 1 apabila tingkat getaran lebih besar dari 4.0, sedangkan suhu pembakaran berlebihan direpresentasikan oleh Pembakaran_Too_High, yang bernilai 1 jika suhu pembakaran melampaui 420°C. Indikasi cacat material ditentukan oleh RFID_Bad berdasarkan pembacaan RFID.

Konsistensi proses termal dinyatakan melalui PressureTempIndex dengan rumus $F_5 = T \times P$. Efisiensi proses direpresentasikan oleh EfficiencyIndex dengan rumus $F_6 = \frac{U}{H}$, sedangkan beban panas total dinyatakan oleh ThermalLoad dengan rumus $F_7 = R \times H$.

Tabel 2. Daftar Fitur Turunan

Fitur	Makna
Delta_Suhu	Selisih suhu yang memicu tekanan termal
HumidityHeat	Beban panas-lembap
MachineStress	Tekanan mekanis
CoolingRate	Indikator perubahan suhu mendadak
Suhu_Rendam_Too_High	Kondisi rendaman ekstrem
Getaran_Ekstrim	Getaran berbahaya
Pembakaran_Too_High	Suhu pembakaran berlebihan
RFID_Bad	Indikasi cacat material
PressureTempIndex	Konsistensi proses termal
EfficiencyIndex	Efisiensi proses
ThermalLoad	Beban panas total

2.5 Preprocessing Data

Tahapan preprocessing meliputi:

1. Seleksi dan penyusunan fitur, dengan mengintegrasikan fitur proses dan fitur turunan yang relevan untuk membentuk dataset pelatihan yang representatif. Pembersihan struktur data, seperti penamaan kolom, tipe data, dan pengecekan jumlah baris.
2. Penanganan kualitas data, termasuk pemeriksaan nilai tidak valid, ketidakkonsistenan data, dan penyesuaian format data agar sesuai dengan kebutuhan pemodelan.
3. Pemilihan fitur, di mana hanya fitur rekayasa digunakan dalam proses pelatihan model karena memiliki pola yang lebih stabil dibanding fitur mentah.

Tidak dilakukan normalisasi karena algoritma Random Forest tidak sensitif terhadap skala data.

2.6 Pembentukan Label

Label pada penelitian ini ditentukan berdasarkan catatan historis kegagalan chip RFID dari PT Nusa Solar Indonesia. Pada kondisi data kegagalan tidak lengkap, pendekatan risk scoring digunakan sebagai pendukung untuk memvalidasi konsistensi label. Chip diklasifikasikan sebagai 0 (Aman) dan 1 (Rusak) berdasarkan status kegagalan aktual, sehingga proses pelatihan model tidak mengalami data leakage.

2.7 Penyeimbangan Data SMOTE

Dataset yang digunakan dalam penelitian ini menunjukkan ketidakseimbangan kelas, di mana jumlah chip berlabel 0 (Aman) lebih besar dibandingkan chip berlabel 1 (Rusak), dengan rasio kelas sekitar 2:1. Kondisi ini berpotensi menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas dan menurunkan kemampuan deteksi kelas minoritas.



Untuk mengatasi permasalahan tersebut, digunakan metode Synthetic Minority Over-sampling Technique (SMOTE). Penerapan SMOTE dilakukan setelah data dibagi menjadi data latih dan data uji, dan hanya diterapkan pada data latih, guna menghindari terjadinya data leakage dan overfitting. Melalui proses ini, jumlah sampel pada kelas minoritas ditingkatkan hingga distribusi kelas pada data latih menjadi seimbang dengan rasio 1:1.

Penyeimbangan data ini bertujuan untuk memastikan bahwa model Random Forest memperoleh representasi yang seimbang dari kedua kelas selama proses pelatihan, sehingga kinerja model dalam mendeteksi chip RFID yang mengalami kerusakan dapat ditingkatkan.

2.8 Model (Random Forest Classifier)

Algoritma Random Forest digunakan sebagai model utama dalam penelitian ini karena memiliki kemampuan yang baik dalam menangani hubungan non-linear antarvariabel, menunjukkan performa yang stabil pada dataset berdimensi tinggi, serta relatif tidak mudah mengalami overfitting karena berbasis pada mekanisme ensemble dari banyak pohon keputusan, sekaligus menyediakan interpretabilitas melalui analisis feature importance. Dalam implementasinya, model ini dikonfigurasi dengan parameter $n_estimators$ sebesar 300 untuk memastikan jumlah pohon yang cukup dalam membangun prediksi yang robust, max_depth dibiarkan tanpa batasan agar setiap pohon dapat tumbuh optimal sesuai kompleksitas data, serta $random_state$ diset pada nilai 42 untuk menjamin reproduktibilitas hasil. Model kemudian dilatih menggunakan dataset yang telah melewati tahapan preprocessing, feature engineering, dan penyeimbangan data sehingga kualitas dan representativitas data dapat mendukung kinerja model secara maksimal.

2.9 Evaluasi Model

Penilaian kinerja model menggunakan beberapa metrik evaluasi, yaitu Accuracy, Precision, Recall, F1-Score, ROC, dan AUC, sebagaimana dirangkum pada Tabel 3.

- a. Accuracy digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data dan dirumuskan sebagai

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- b. Precision mengukur tingkat ketepatan prediksi positif dan didefinisikan sebagai

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

- c. Recall mengukur kemampuan model dalam mengidentifikasi kelas positif dan dirumuskan sebagai

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

- d. F1-Score merepresentasikan keseimbangan precision dan recall, dirumuskan

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

- e. Receiver Operating Characteristic (ROC) digunakan untuk mengevaluasi kinerja model berdasarkan true positive rate dan false positive rate. Nilai true positive rate dirumuskan sebagai

$$TPR = \frac{TP}{TP+FN} \quad (5)$$

sedangkan false positive rate dirumuskan sebagai

$$FPR = \frac{FP}{FP+TN} \quad (6)$$

- f. Area Under the Curve (AUC) digunakan untuk mengukur kemampuan diskriminasi model secara keseluruhan dan didefinisikan sebagai luas area di bawah kurva ROC, yang dirumuskan sebagai.

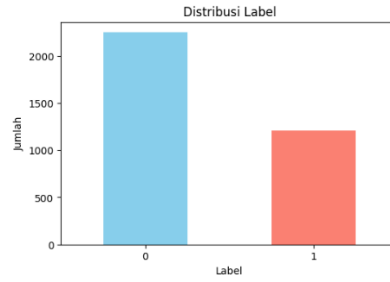
$$AUC = \int_0^1 TPR(FPR) d(FPR). \quad (7)$$

Metrik ini dipilih karena mampu merepresentasikan kinerja model dalam mengidentifikasi kasus ‘Rusak’, yang memiliki prioritas lebih tinggi dalam konteks manufaktur.

3. HASIL DAN PEMBAHASAN

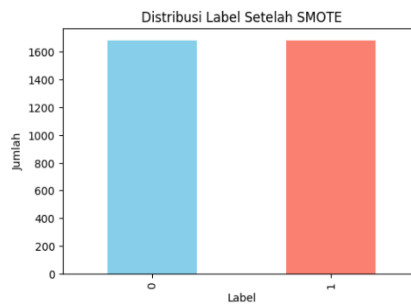
3.1 Analisis Awal Data dan Teori Pendukung

Analisis awal dilakukan untuk memahami karakteristik dataset sebelum pelatihan model. Distribusi label pada Gambar 3 menunjukkan bahwa jumlah chip Aman (0) jauh lebih banyak dari pada Rusak (1). Ketidakseimbangan ini merupakan hal yang umum pada data industri dan berpotensi menyebabkan model bias terhadap kelas mayoritas.



Gambar 3. Distribusi Label Sebelum SMOTE

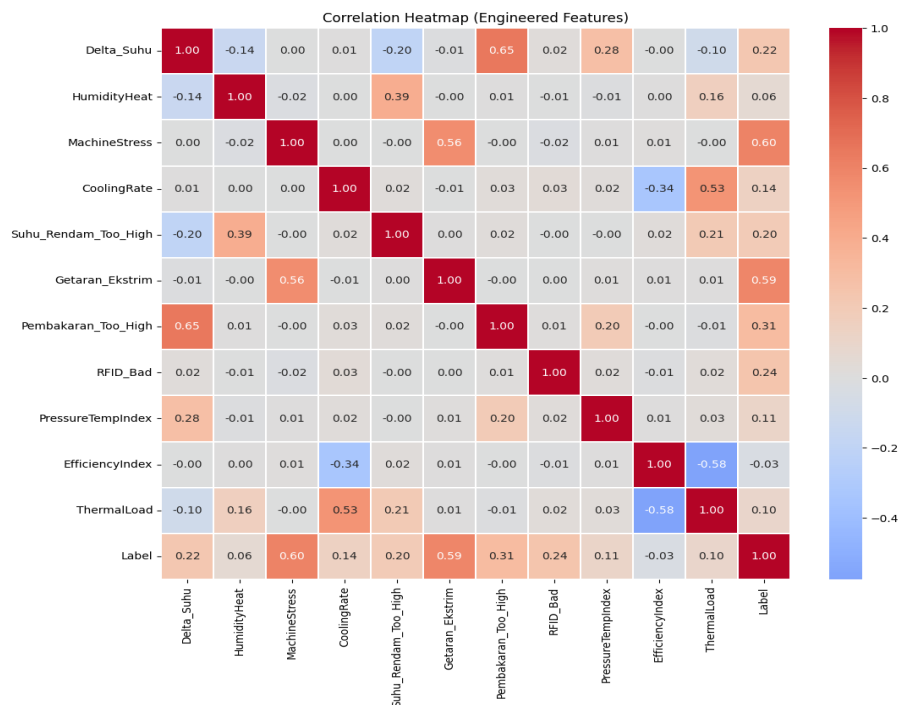
Untuk mengatasi ketidakseimbangan tersebut, digunakan teknik SMOTE untuk menyeimbangkan jumlah sampel pada kelas minoritas. Distribusi label setelah penerapan SMOTE ditunjukkan pada Gambar 4, di mana jumlah sampel pada kelas Aman dan Rusak terlihat seimbang. Keseimbangan distribusi ini membantu model memahami karakteristik kerusakan dengan lebih baik serta mengurangi kemungkinan kesalahan prediksi, khususnya false negative.



Gambar 4. Distribusi Label Setelah SMOTE

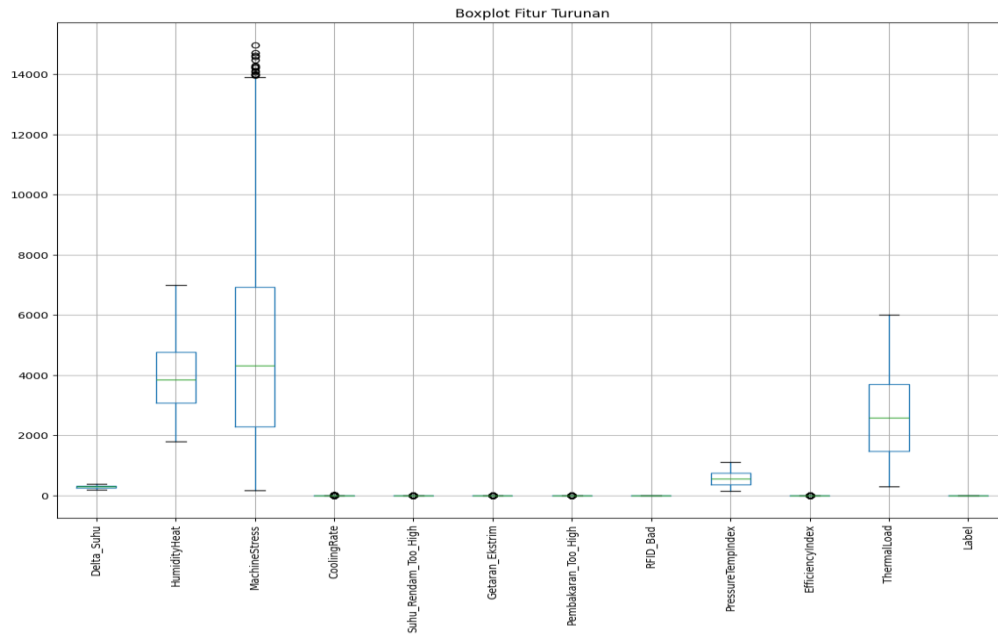
3.2 Analisis Hubungan Fitur Turunan

Gambar 5 menunjukkan matriks korelasi antar fitur turunan, di mana fitur MachineStress, ThermalLoad, dan PressureTempIndex memiliki korelasi yang relatif kuat terhadap label kerusakan chip. Hal ini menunjukkan bahwa tekanan mekanis dan beban termal, baik secara individual maupun melalui interaksinya, berperan penting dalam meningkatkan risiko kerusakan chip RFID. Temuan ini mengindikasikan bahwa kerusakan chip tidak dipengaruhi oleh satu parameter tunggal, melainkan oleh kombinasi kondisi proses, sehingga penggunaan fitur turunan menjadi relevan untuk merepresentasikan hubungan antar-parameter tersebut.



Gambar 5. Kolerasi Heatmap Fitur Turunan

Boxplot pada Gambar 6 memperlihatkan sebaran nilai fitur turunan tersebut, di mana keberadaan outlier menunjukkan adanya variasi ekstrem pada kondisi proses yang merupakan karakteristik alami data industri.



Gambar 6. Boxplot Fitur Turunan

3.3 Hasil Pelatihan Model

Model Random Forest digunakan karena kemampuannya menangani hubungan non-linear antar variabel serta memberikan interpretasi melalui feature importance. Tabel 4 menunjukkan metrik evaluasi utama, di mana model menghasilkan nilai precision, recall, dan f1-score yang baik. Ada kelas Rusak, nilai recall yang sangat tinggi (0,967) menandakan bahwa model mampu mendeteksi hampir seluruh chip yang benar-benar mengalami kerusakan, sehingga risiko false negative dapat diminimalkan. Namun, nilai precision pada kelas ini sebesar 0,844 menunjukkan masih adanya sebagian prediksi chip rusak yang sebenarnya berada dalam kondisi aman (false positive), yang mencerminkan adanya trade-off antara sensitivitas dan ketepatan prediksi. Sementara itu, pada kelas Aman, model menghasilkan nilai precision dan recall yang tinggi, menunjukkan konsistensi dalam mengklasifikasikan chip yang tidak rusak. Secara keseluruhan, keseimbangan nilai F1-score serta macro dan weighted average yang relatif stabil mengindikasikan bahwa model tidak bias terhadap salah satu kelas dan layak digunakan sebagai sistem peringatan dini untuk mendeteksi potensi kerusakan chip RFID pada proses manufaktur sel surya.

Tabel 3. Hasil Classification Report Model RF

	Precision	Recall	F1-score	Support
Aman	0.9807	0.9039	0.9407	562.0000
Rusak	0.8444	0.9670	0.9015	303.0000
accuracy	0.9260	0.9260	0.9260	0.9260
macro avg	0.9125	0.9355	0.9211	865.0000
weighted avg	0.9329	0.9260	0.9270	865.0000

Confusion matrix pada Tabel 5 menunjukkan bahwa sebagian besar prediksi model berada pada klasifikasi yang tepat. Dari total data, True Negative mencapai sekitar 58,7% dan True Positive sebesar 33,9%, yang menandakan bahwa model mampu mengklasifikasikan chip aman maupun rusak dengan baik. Proporsi False Negative yang rendah, yaitu sekitar 1,2%, menunjukkan bahwa hanya sedikit chip yang sebenarnya rusak namun diprediksi sebagai aman, sehingga risiko gangguan pada proses produksi dapat diminimalkan. Sementara itu, False Positive sebesar sekitar 6,2% menunjukkan adanya chip yang sebenarnya masih aman tetapi diprediksi rusak, yang berdampak pada perlunya pemeriksaan tambahan atau verifikasi lanjutan sebelum chip melanjutkan ke tahap produksi berikutnya. Meskipun kondisi ini dapat menambah beban pemeriksaan, pola kesalahan tersebut relatif lebih dapat ditoleransi dibandingkan False Negative karena tidak menimbulkan kegagalan proses produksi secara langsung.

Tabel 4. Confusion Matrix Hasil Prediksi Model

	Pred_0	Pred_1
Actual_0	508 (58.7%)	54 (6.2%)

	Pred_0	Pred_1
Actual 1	10 (1.2%)	293 (33.9%)

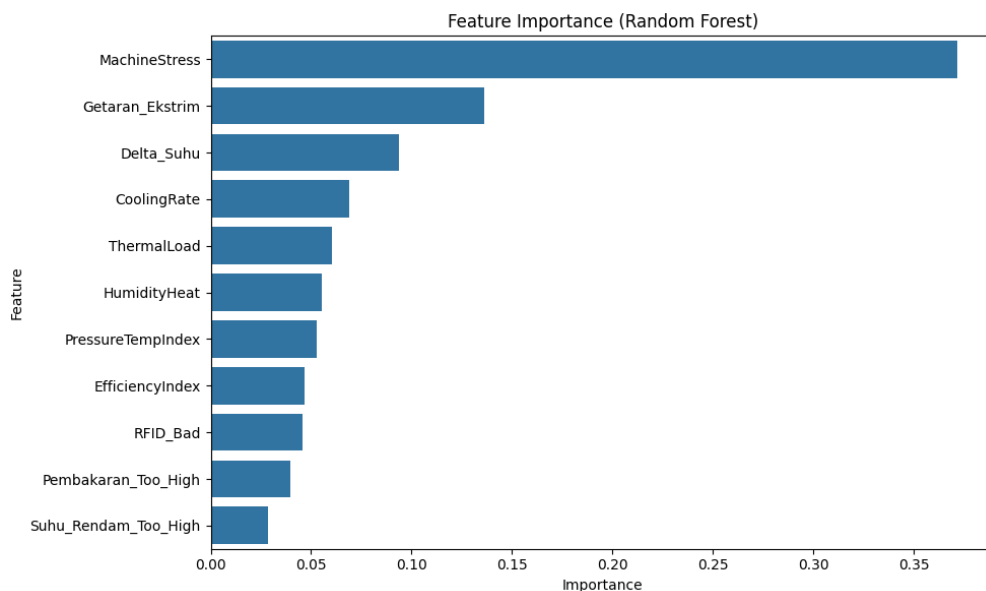
3.4 Analisis Feature Importance dan Evaluasi Tambahan

Tabel 6 dan Gambar 7 menunjukkan kontribusi setiap fitur turunan terhadap keputusan model. Fitur seperti MachineStress dan ThermalLoad memberikan pengaruh paling besar terhadap prediksi kerusakan. Hal ini memperkuat pemahaman bahwa faktor mekanis dan termal sangat relevan dalam konteks kualitas chip RFID. Fitur lain seperti Delta_Suhu, EfficiencyIndex, dan CoolingRate memberikan kontribusi menengah dan tetap berperan dalam mendukung keputusan model. Seluruh fitur yang digunakan merupakan parameter proses yang tersedia sebelum tahap inspeksi akhir, sehingga model prediksi terhindar dari potensi data leakage. Visualisasi feature importance pada Gambar 7 memperkuat hasil pada Tabel 6, di mana MachineStress, ThermalLoad, dan PressureTempIndex tetap menjadi kontributor utama.

Temuan bahwa MachineStress dan ThermalLoad merupakan fitur paling penting dalam memprediksi kerusakan chip RFID selaras dengan arah penelitian terkini yang mengeksplorasi penggunaan RFID dalam monitoring kondisi dan maintenance berbasis machine learning. Abdul Sami et al. menyajikan ulasan komprehensif mengenai kemajuan sensor RFID pasif untuk aplikasi Structural Health Monitoring, termasuk metode ekstraksi fitur dan tantangan teknis sensor yang relevan dengan deteksi kondisi struktur dan kegagalan komponen secara real-time, yang memperkuat relevansi fitur fisiologis yang digunakan dalam penelitian ini[26]. Selain itu, penelitian sistematis di bidang predictive maintenance berbasis AI dan machine learning menegaskan bahwa penyertaan data kondisi nyata dan algoritma prediktif sangat penting dalam meningkatkan akurasi prediksi kerusakan dan deteksi dini di berbagai aplikasi industri[27]. Studi lain melaporkan desain dan implementasi sensor RFID chipless inovatif untuk deteksi retakan dan perubahan struktural, yang menunjukkan tren peningkatan integrasi RFID dengan teknik machine learning untuk aplikasi monitoring kondisi dan prediksi kegagalan di lingkungan keras[28].

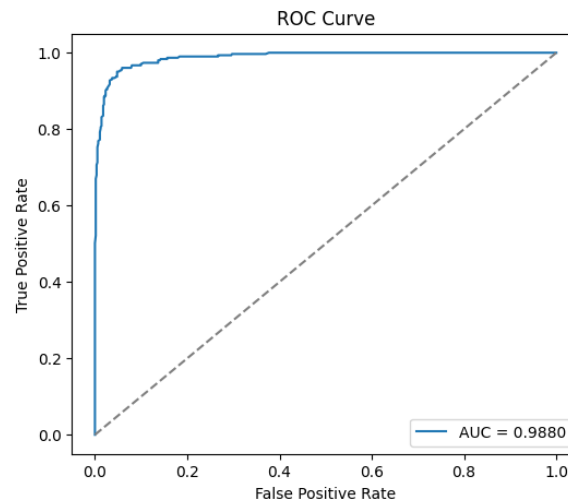
Tabel 5. Peringkat Feature Importance Model

Peringkat	Fitur	Importance	Peringkat	Fitur	Importance
1	MachineStress	0.29	7	HumidityHeat	0.03
2	ThermalLoad	0.21	8	Pembakaran_Too_High	0.02
3	PressureTempIndex	0.18	9	Getaran_Ekstrim	0.02
4	Delta_Suhu	0.11	10	Suhu_Rendam_Too_High	0.01
5	EfficiencyIndex	0.08			
6	CoolingRate	0.05			



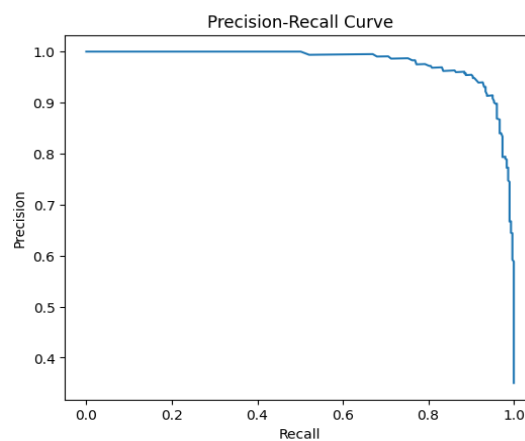
Gambar 7. Visualisasi Feature Importance Model

Evaluasi kinerja model dilakukan menggunakan ROC Curve pada Gambar 8, yang menunjukkan kemampuan model dalam membedakan kelas positif dan negatif pada berbagai nilai ambang (threshold). Berdasarkan hasil evaluasi, model menghasilkan nilai AUC sebesar 0,988, yang mengindikasikan kemampuan diskriminasi yang sangat baik dalam memisahkan chip dalam kondisi aman dan rusak. Nilai AUC yang mendekati 1 menunjukkan bahwa model mampu mencapai True Positive Rate yang tinggi pada False Positive Rate yang rendah, sehingga performa klasifikasi tetap konsisten pada berbagai threshold.



Gambar 8. ROC Curve Model Random Forest

Selanjutnya, Precision–Recall Curve pada Gambar 9 menunjukkan bahwa nilai precision berada mendekati 1,0 pada sebagian besar rentang recall. Precision tetap tinggi hingga recall mencapai sekitar 0,8–0,9, menandakan kemampuan model mendeteksi chip rusak dengan ketepatan yang baik. Penurunan precision mulai terlihat ketika recall mendekati nilai maksimum ($\approx 1,0$), mencerminkan adanya trade-off antara ketepatan prediksi dan kemampuan deteksi. Secara keseluruhan, karakteristik ini menunjukkan bahwa model mampu menjaga performa klasifikasi yang stabil dan layak digunakan sebagai sistem peringatan dini.



Gambar 9. Precision-Recall Curve Model Random Forest

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa kerusakan chip RFID pada proses produksi sel surya dapat diprediksi dengan baik menggunakan model Random Forest ketika data dilengkapi dengan fitur turunan yang merepresentasikan kondisi fisik proses, seperti beban mekanis, perubahan suhu, dan kombinasi tekanan–termal. Hasil penelitian menunjukkan bahwa fitur-fitur tersebut mampu menggambarkan pola kerusakan secara lebih akurat dibandingkan data mentah, sehingga model memiliki kemampuan yang lebih stabil dalam membedakan chip yang aman dan yang berpotensi rusak. Penerapan teknik penyeimbangan data seperti SMOTE turut berperan penting dalam meningkatkan sensitivitas model terhadap kasus kerusakan yang jumlahnya relatif sedikit pada data asli. Secara keseluruhan, penelitian ini menjawab bahwa pendekatan machine learning berbasis fitur turunan dapat dijadikan alat bantu deteksi dini untuk meminimalkan risiko gangguan produksi dan meningkatkan keandalan sistem pelacakan material di lingkungan manufaktur sel surya.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada dosen pembimbing yang telah memberikan bimbingan, arahan, dan masukan berharga selama proses penelitian dan penyusunan manuskrip ini. Ucapan terima kasih juga disampaikan kepada PT Nusa Solar Indonesia selaku tempat pelaksanaan penelitian yang telah menyediakan data, fasilitas, dan dukungan teknis yang memungkinkan penelitian ini dapat dilaksanakan dengan baik. Selain itu, penulis



menyampaikan penghargaan kepada Universitas Internasional Batam sebagai institusi afiliasi yang telah memberikan dukungan akademik dan administratif sehingga penelitian ini dapat terselesaikan dengan baik.

REFERENCES

- [1] S. Butsianto and S. Herdiyana, "SISTEM INVENTORI BERBASIS INTERNET OF THINGS MENGGUNAKAN RADIO FREQUENCY IDENTIFICATION (STUDI KASUS PT. GRAFITECINDO CIPTAPRIMA)," *JSI : Jurnal Sistem Informasi (E-Journal)*, vol. 13, no. 2, pp. 2298–2314, 2021, doi: <https://doi.org/10.18495/jsi.v13i2.47>.
- [2] E. PUICA, "Product Personalization and Customization: Proposing a System Architecture that Integrates Self-Transactional Materials with RFID and IoT Shared Database," *Informatica Economica*, vol. 27, no. 3/2023, pp. 5–16, Sep. 2023, doi: 10.24818/issn14531305/27.3.2023.01.
- [3] A. Budiayanto and M. Muslim, "Sinergi International Journal of Logistics Optimizing Inventory Systems with RFID: A Narrative Review of Integration, Efficiency, and Barriers," *Sinergi International Journal of Logistics*, vol. 2, no. 2, pp. 133–146, 2024, Accessed: Jan. 02, 2026. [Online]. Available: https://journal.sinergi.or.id/index.php/ijl/article/view/621?utm_source=chatgpt.com
- [4] J. Ferdousmou, M. Prabha, M. O. Farouk, M. Samiun, H. M. Sozib, and A. M. Zaman, "IoT-Enabled RFID in Supply Chain Management: A Comprehensive Survey and Future Directions," *Journal of Computer and Communications*, vol. 12, no. 11, pp. 207–223, 2024, doi: 10.4236/jcc.2024.1211015.
- [5] S. Yang, X. Wang, Q. Song, X. Liu, P. Li, and Y. Yang, "Passive Strain Sensor Based on RFID Patch Antenna," in *Journal of Physics: Conference Series*, Institute of Physics, 2025. doi: 10.1088/1742-6596/3056/1/012015.
- [6] R. A. F. Fattah, and A. R. Manga, "Karya Ilmiah Makalah Alat Sistem Pintu Kunci Otomatis Pada Ruang Menggunakan Sensor RFID," *LINIER: Literatur Informatika dan Komputer*, vol. 2, no. 1, pp. 109–114, Apr. 2025, doi: 10.33096/linier.v2i1.2792.
- [7] F. A. Rodríguez-Corbo et al., "Deterministic 3D ray-launching millimeter wave channel characterization for vehicular communications in urban environments," *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–25, Sep. 2020, doi: 10.3390/s20185284.
- [8] M. Peggior, R. Arablouei, M. Nikzad, and S. Masood, "Displacement Estimation via 3D-Printed RFID Sensors for Structural Health Monitoring: Leveraging Machine Learning and Photoluminescence to Overcome Data Gaps," *Sensors*, vol. 24, no. 4, Feb. 2024, doi: 10.3390/s24041233.
- [9] M. Kilibarda, M. Andrejic, and V. Pajic, "Supply Chain RFID Solution Evaluation Applying AHP and FAHP Methods: A Case Study of the Serbian Market," *Tehnicki Vjesnik*, vol. 29, no. 6, pp. 1811–1818, 2022, doi: 10.17559/TV-20210330142920.
- [10] A. A. A. Hakeem, D. Solyali, M. Asmael, and Q. Zeeshan, "Smart Manufacturing for Industry 4.0 using Radio Frequency Identification (RFID) Technology," *Jurnal Kejuruteraan*, vol. 32, no. 1, pp. 31–38, Feb. 2020, doi: 10.17576/jkukm-2020-32(1)-05.
- [11] S. Bi et al., "A Novel RFID Localization Approach to Smart Self-Service Borrowing and Returning System," *CMES - Computer Modeling in Engineering and Sciences*, vol. 135, no. 1, pp. 527–538, 2023, doi: 10.32604/cmcs.2022.022298.
- [12] Haeruddin, E. Noersasongko, Purwanto, and Muljono, "A Multi-Model Framework for Rainfall Forecasting: Evaluating Performance Model Statistical, Machine Learning, and Deep Learning Methods," in *2025 International Conference on Smart Computing, IoT and Machine Learning, SIML 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/SIML65326.2025.11080798.
- [13] Y. Christian, "Rental Price Prediction of Boarding Houses in Batam City Using Linear Regression and Random Forest Algorithms," *Journal of Applied Informatics and Computing (JAIC)*, vol. 7, no. 2, pp. 263–270, 2023, [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [14] H. Zermane and A. Drardja, "Development of an Efficient Cement Production Monitoring System Based on the Improved Random Forest Algorithm," *The International Journal of Advanced Manufacturing Technology*, vol. 120, no. 3–4, pp. 1853–1866, Sep. 2022, doi: 10.1007/s00170-022-08884-z.
- [15] A. Meddaoui, A. Hachmoud, and M. Hain, "Advanced ML for Predictive Maintenance: Case Study on Remaining Useful Life Prediction and Reliability Enhancement," *The International Journal of Advanced Manufacturing Technology*, vol. 132, pp. 323–335, Jan. 2024, doi: 10.21203/rs.3.rs-3875020/v1.
- [16] M. Etxegarai, M. Camps, L. Echeverria, M. Ribalta, F. Bonada, and X. Domingo, "Virtual Sensors for Smart Data Generation and Processing in AI-Driven Industrial Applications," in *Industry 4.0 — Perspectives and Applications*, IntechOpen, 2022. Accessed: Jan. 02, 2026. [Online]. Available: <https://www.intechopen.com/chapters/106988>
- [17] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 677–690, Jul. 2022, doi: 10.30812/matrik.v21i3.1726.
- [18] G. G. Setiaji, J. Suntoro, and A. Rifa'i, "Random State Parameter Undersampling untuk Penanganan Data dengan Kelas Tidak Seimbang pada Algoritme Random Forest," *Jurnal Transformatika*, vol. 21, no. 2, pp. 73–83, Mar. 2024, doi: 10.26623/transformatika.v21i2.8901.
- [19] V. Arpitha and A. K. Pani, "Machine Learning Approaches for Fault Detection in Semiconductor Manufacturing Process: A Critical Review of Recent Applications and Future Perspectives," *Chem Biochem Eng Q*, vol. 36, no. 1, pp. 1–16, 2022, doi: 10.15255/CABEQ.2021.1973.
- [20] M. P. Utami, F. Suroso, F. Lailasari H., F. P. J. Sibuea, and K. Chandra, "Integrasi Algoritma Support Vector Machine dengan Java untuk Memprediksi Kualitas Komponen Otomotif dalam Industri 4.0," *Techno.Com*, vol. 24, no. 3, pp. 790–797, Aug. 2025, doi: 10.62411/tc.v24i3.12719.
- [21] T. Da Rocha, A. B. Canciglieri, and A. L. Szejka, "Towards a machine learning failure prediction system applied to a smart manufacturing process," in *Product Lifecycle Management Enabling Smart X*, Springer International Publishing, 2020, pp. 26–35. doi: 10.1007/978-3-030-62807-9_3.



- [22] F. Ruliansyah and M. Daud, “Sistem Pendataan Kunjungan pada Perpustakaan Fakultas Teknik UNIMAL Menggunakan Modul RFID Visiting Data Collection System at the Engineering Faculty of UNIMAL Using an RFID Module,” *TELKA*, vol. 8, no. 2, pp. 103–115, 2022.
- [23] J. Liang et al., “Analysis of Electromagnetic Interference for Anti-Medal UHF RFID Temperature Tag in High Power Electronic Equipment,” *Electronics (Switzerland)*, vol. 12, no. 17, Sep. 2023, doi: 10.3390/electronics12173577.
- [24] T. Xu, H. Wang, L. Feng, and Y. Zhu, “Risk Factors Assessment of Smart Supply Chain in Intelligent Manufacturing Services Using DEMATEL Method With Linguistic q-ROF Information,” *Journal of Operations Intelligence*, vol. 2, no. 1, pp. 129–152, Nov. 2024, doi: 10.31181/jopi21202417.
- [25] C. Herdian, A. Kamila, F. Feiters Tampinongkol, A. S. Kembau, G. Agung, and M. Budidarma, “ONE-HOT ENCODING FEATURE ENGINEERING UNTUK LABEL-BASED DATA STUDI KASUS PREDIKSI HARGA MOBIL BEKAS,” *Informasi Interaktif: Jurnal Informatika dan Teknologi Informasi*, vol. 9, no. 1, pp. 10–16, 2021, doi: <https://doi.org/10.37159/jii.v9i1.41>.
- [26] A. Sami, G. Yun Tian, A. Mahmud Jaya Marindra, and H. Song, “Journal Pre-proof A Comprehensive Review on Passive RFID Sensors for Structural Health Monitoring: Developments, Challenges, and Opportunities A Comprehensive Review on Passive RFID Sensors for Structural Health Monitoring: Developments, Challenges, and Opportunities,” *Chinese Journal of Mechanical Engineering Song*, vol. 38, no. 1, pp. 1–22, 2025, doi: 10.1016/j.cjme.2025.100117.
- [27] L. Rojas, Á. Peña, and J. Garcia, “AI-Driven Predictive Maintenance in Mining: A Systematic Literature Review on Fault Detection, Digital Twins, and Intelligent Asset Management,” *Applied Sciences (Switzerland)*, vol. 15, no. 6, p. 3337, Mar. 2025, doi: 10.3390/app15063337.
- [28] B. Wang, R. Hou, K. Wang, S. Cao, C. Yuan, and M. Feng, “Passive RFID sensor design and research for metal crack width detection,” *Results in Engineering*, vol. 26, no. 7, Jun. 2025, doi: 10.1016/j.rineng.2025.105637.