

**JEMSI:**  
**Jurnal Ekonomi Manajemen Sistem**  
**Informasi**

E-ISSN: 2686-5238  
P-ISSN: 2686-4916

<https://dinastirev.org/JEMSI> [dinasti.info@gmail.com](mailto:dinasti.info@gmail.com) [+62 811 7404 455](tel:+628117404455)

DOI: <https://doi.org/10.38035/jemsi.v6i6>  
<https://creativecommons.org/licenses/by/4.0/>

## Klasifikasi Barang NG Berdasarkan Dataset Menggunakan Algoritma C.45 Studi Kasus: Perusahaan Manufaktur XYZ

Jasmine Dina Sabila<sup>1</sup>, Tukino Tukino<sup>2</sup>, Agustia Hananto<sup>3</sup>

<sup>1</sup>Universitas Buana Perjuangan Karawang, Karawang, Indonesia, [si22.jasminesabila@mhs.ubpkarawang.ac.id](mailto:si22.jasminesabila@mhs.ubpkarawang.ac.id)

<sup>2</sup>Universitas Buana Perjuangan Karawang, Karawang, Indonesia, [tukino@ubpkarawang.ac.id](mailto:tukino@ubpkarawang.ac.id)

<sup>3</sup>Universitas Buana Perjuangan Karawang, Karawang, Indonesia, [agustia.hananto@ubpkarawang.ac.id](mailto:agustia.hananto@ubpkarawang.ac.id)

Corresponding Author: [si22.jasminesabila@mhs.ubpkarawang.ac.id](mailto:si22.jasminesabila@mhs.ubpkarawang.ac.id)<sup>1</sup>

**Abstract:** *Company XYZ is a Japanese company engaged in the automotive sector, especially in the production of spare parts. Its main product is air-conditioning hose assemblies for Original Equipment Manufacturer (OEM) vehicles. The Final Inspection production department is tasked with ensuring that production goods meet quality standards in accordance with the Work Instruction (WI) and dimensional specifications based on gauges. In carrying out its duties, if Not Good (NG) is found, the goods will be separated from OK goods, then marked and set aside for further analysis. Recording of Not Good (NG) goods is done through the Worktop Website, where data can be accessed at any time and includes the type, quantity, and NG (not good) problems. This system helps real-time reporting and facilitates monitoring of production data. From the results of the logbook input, it was found that there were repeated NG (not good) in the production process. This recurring NG condition disrupts the cycle time and impacts production efficiency. The unprocessed logbook data can be transformed into knowledge to support decision-making in the production department by classifying NG data into severity categories using data mining techniques. This classification is expected to identify patterns that help in making better decisions.*

**Keyword:** *Website Worktop, Data Mining, Not Good (NG)*

**Abstrak:** Perusahaan XYZ adalah perusahaan asal Jepang yang bergerak di bidang otomotif, khususnya dalam produksi suku cadang. Produk utamanya adalah perakitan selang untuk sistem pendingin udara (air-conditioning hose assemblies) untuk kendaraan Original Equipment Manufacturer (OEM). Departemen produksi Final Inspection bertugas memastikan bahwa barang produksi memenuhi standar kualitas sesuai dengan Work Instruction (WI) dan spesifikasi dimensi berdasarkan gauge. Dalam pelaksanaan tugasnya, apabila ditemukan Not Good (NG), barang tersebut akan dipisahkan dari barang OK, kemudian ditandai dan disisihkan untuk dianalisis lebih lanjut. Pencatatan barang Not Good (NG) ini dilakukan melalui Website Worktop, Dimana data dapat diakses kapan saja dan mencakup jenis, jumlah, serta masalah NG (*not good*). Sistem ini membantu pelaporan real-time dan mempermudah pemantauan data produksi. Dari hasil input logbook, ditemukan adanya NG (*not good*) yang berulang pada proses produksi. Kondisi NG (*not good*) yang berulang ini menyebabkan gangguan pada waktu cycle time, sehingga berdampak pada efisiensi produksi. Data hasil input logbook yang belum

diolah menjadi knowledge dapat membantu membuat keputusan departemen produksi dengan mengklasifikasikan data NG (*not good*) menjadi kategori keparahan NG (*not good*) menggunakan data mining diharapkan dapat membentuk pola untuk mendukung pengambilan keputusan yang baik.

**Kata Kunci:** *Website Worktop, Data Mining, Not Good (NG)*

---

## PENDAHULUAN

Dalam era manufaktur modern, pengendalian kualitas produk menjadi aspek krusial yang menentukan daya saing perusahaan dalam industri otomotif. Pengembangan teknologi informasi dan komputasi telah memungkinkan pemanfaatan data historis produksi untuk mengoptimalkan proses identifikasi dan klasifikasi produk cacat (Not Good/NG) secara lebih sistematis dan terukur (Martha & Anggraini, 2021). Analisis terhadap logbook produksi mengindikasikan adanya pola kemunculan produk NG (*not good*) yang berulang, fenomena ini mendorong kebutuhan akan sistem klasifikasi yang dapat mengkategorikan data NG (*not good*) berdasarkan tingkat keparahan dengan memanfaatkan pendekatan data mining untuk menghasilkan pola yang mendukung pengambilan keputusan yang lebih efektif (Setyani & Gunadi, 2020).

Data mining sebagai teknik ekstraksi pengetahuan dari kumpulan data yang besar memiliki lima fungsi dasar, dan salah satunya adalah klasifikasi yang dapat dimanfaatkan untuk mengidentifikasi pola signifikan serta memprediksi tren data di masa mendatang. Kajian terdahulu yang dilakukan oleh (Martha & Anggraini, 2021).

Algoritma C.45 merupakan metode klasifikasi yang menghasilkan pohon keputusan sebagai keluaran dengan memanfaatkan sampel pelatihan (training samples) sebagai data contoh yang telah terverifikasi kebenarannya. Metodologi ini bekerja dengan mencari atribut yang paling informatif untuk memisahkan data ke dalam berbagai kelas melalui perhitungan entropy dan gain sebagai parameter evaluasi efektivitas variabel dalam klasifikasi data (Nurdiyani Sari & Azharina, 2024).

Teknologi Informasi, khususnya sebagai analisis data, yang meliputi pemrograman dan scripting, statistik dan matematika, data mining, dan cloud computing. Kompetensi-kompetensi ini sangat relevan dengan penelitian ini yang mengimplementasikan algoritma C.45 untuk klasifikasi data NG (*not good*) dengan memanfaatkan berbagai tools dan teknik analisis data (Purwanto et al., 2023). Fungsi utama sistem operasi adalah menyederhanakan kompleksitas teknis dan memberikan kemudahan bagi programmer dalam bekerja dengan instruksi-instruksi yang digunakan, yang sangat relevan dengan implementasi algoritma C.45.

Penelitian ini bertujuan untuk mengklasifikasikan data NG (*not good*) menggunakan algoritma C.45 dan mengevaluasi tingkat akurasi klasifikasi tersebut. Hasil penelitian diharapkan dapat memberikan kontribusi signifikan bagi perusahaan manufaktur XYZ, dalam mengidentifikasi pola dan faktor-faktor yang berkontribusi terhadap kemunculan produk NG (*not good*), dan dapat memperkaya literatur akademik tentang penerapan algoritma C.45 dalam klasifikasi data NG (*not good*) di industri manufaktur, khususnya dalam konteks perakitan selang sistem pendingin udara kendaraan.

## METODE

Penelitian ini dilaksanakan di perusahaan manufaktur XYZ. Pengumpulan data dilaksanakan pada Departemen Final Inspection yang bertugas memastikan kualitas produk sebelum pengiriman ke pelanggan. Populasi penelitian mencakup seluruh data produk Not Good (NG) yang tercatat melalui sistem Website Worktop pada rentang waktu penelitian. Data diperoleh melalui observasi langsung terhadap proses inspeksi dan ekstraksi dari database

perusahaan yang didokumentasikan dalam format Excel. Proses identifikasi NG melibatkan pemeriksaan visual dan pengukuran dimensi menggunakan alat ukur gauge dengan tingkat presisi tinggi, di mana produk diklasifikasikan berdasarkan kesesuaian terhadap standar kualitas Perusahaan.

Instrumentasi penelitian meliputi alat ukur gauge untuk verifikasi dimensi, komputer dengan aplikasi Website Worktop untuk dokumentasi dan algoritma C.45 yang diimplementasikan untuk klasifikasi data. Pengolahan data menggunakan metodologi Knowledge Discovery in Database (KDD) yang mencakup tahapan seleksi data, pra-pemrosesan dengan penghapusan duplikasi dan koreksi kesalahan, transformasi data, implementasi algoritma C.45 untuk klasifikasi, dan evaluasi hasil. Algoritma C.45 diterapkan dengan perhitungan entropy dan gain untuk mengukur efektivitas variabel dalam klasifikasi data NG. Validasi model dilakukan melalui pengujian akurasi dengan membandingkan hasil klasifikasi algoritma dengan data aktual. Analisis lanjutan terhadap hasil klasifikasi dilakukan untuk mengidentifikasi faktor-faktor utama yang mempengaruhi kemunculan produk NG dan mengembangkan strategi pengendalian kualitas yang lebih efektif.

## **HASIL DAN PEMBAHASAN**

### **Hasil**

#### **Mining dalam Analisis Kualitas Produksi**

Data mining merupakan proses ekstraksi pengetahuan dari kumpulan data besar dengan tujuan menemukan pola tersembunyi yang bermakna. Menurut (Firdaus, 2020), data mining menggunakan algoritma komputasi untuk mengelompokkan data, mengidentifikasi relasi, dan mengenali pola dari basis data kompleks yang tidak terdeteksi melalui metode analisis konvensional. Dalam konteks industri manufaktur, implementasi data mining berperan vital dalam analisis kualitas produksi melalui identifikasi faktor-faktor yang berkontribusi terhadap produk cacat NG (*not good*) serta pengembangan model prediktif untuk mencegah kemunculannya. (Putra & Furqan, 2024) mengemukakan bahwa integrasi data mining dalam proses pengendalian kualitas mampu meningkatkan efisiensi deteksi anomali produksi hingga 42% dibandingkan metode inspeksi manual, sehingga meminimalisir ketergantungan pada pengalaman subjektif inspector.

Knowledge Discovery in Database (KDD) sebagai kerangka metodologis data mining mengintegrasikan tahapan seleksi data, pra-pemrosesan, transformasi, mining, dan interpretasi hasil untuk menghasilkan wawasan komprehensif dari data produksi. Aplikasi data mining dalam analisis kualitas produk memungkinkan pelacakan pola sistematis dari cacat produksi dan identifikasi akar permasalahannya. Implementasi data mining untuk klasifikasi produk cacat pada industri otomotif menghasilkan peningkatan signifikan dalam efektivitas tindakan korektif hingga 37,5% dibandingkan analisis konvensional. Temuan ini mengkonfirmasi bahwa pendekatan data mining tidak hanya berperan dalam identifikasi produk cacat secara retrospektif, tetapi juga dalam pengembangan sistem preventif yang dapat memprediksi potensi kemunculan cacat produksi berdasarkan parameter-parameter proses yang teridentifikasi melalui analisis data historis.

#### **Algoritma C.45 untuk Klasifikasi Data Produksi**

Algoritma C.45 merupakan pengembangan dari algoritma ID3 yang didesain untuk membangun pohon keputusan melalui analisis rekursif terhadap data sampel dengan menggunakan pendekatan berbasis entropi. Sebagaimana dipaparkan oleh (Saputra & Qoiriah, 2022), algoritma C.45 menggunakan kalkulasi information gain sebagai metrik untuk menyeleksi atribut optimal yang mampu memisahkan data ke dalam klasifikasi berbeda dengan tingkat homogenitas tertinggi.

Implementasi algoritma C.45 dalam klasifikasi produk Not Good (NG) melibatkan proses perhitungan entropi untuk mengukur heterogenitas data dan information gain untuk mengevaluasi kontribusi informatif setiap atribut. Sebagaimana dijelaskan oleh (Setio et al., 2020), algoritma ini bekerja dengan membagi dataset secara rekursif berdasarkan atribut yang memiliki information gain tertinggi, membentuk struktur hierarkis yang merepresentasikan aturan klasifikasi. Penerapan algoritma C.45 dalam klasifikasi cacat produksi pada industri elektronik menunjukkan bahwa model yang dihasilkan tidak hanya memiliki akurasi tinggi (94,2%) tetapi juga interpretabilitas superior yang memungkinkan engineer produksi mengidentifikasi dengan jelas rangkaian kondisi yang mengarah pada kemunculan produk cacat. Kemampuan algoritma C.45 dalam menghasilkan model klasifikasi yang komprehensif dan mudah diinterpretasi menjadikannya pilihan optimal untuk analisis data produksi dalam rangka peningkatan kualitas berkelanjutan.

### **Sistem Pengendalian Kualitas Berbasis Teknologi Informasi**

Sistem pengendalian kualitas berbasis teknologi informasi mengintegrasikan perangkat lunak, basis data, dan algoritma analitik untuk memfasilitasi monitoring, evaluasi, dan peningkatan aspek kualitas produk secara komprehensif. Menurut (Wijaya, 2023), implementasi sistem informasi dalam pengendalian kualitas mampu mentransformasi proses inspeksi dari aktivitas reaktif menjadi proaktif melalui analisis real-time terhadap parameter produksi dan ketidaksesuaian produk. Konfigurasi sistem yang optimal mencakup modul akuisisi data otomatis, pemrosesan dan analisis data terintegrasi, visualisasi hasil, serta mekanisme notifikasi anomali (Efendi et al., 2023). Implementasi sistem pengendalian kualitas berbasis data pada industri manufaktur komponen otomotif berhasil mereduksi tingkat produk cacat hingga 27,3% dalam periode implementasi enam bulan, mengonfirmasi signifikansi sistem informasi dalam optimalisasi proses produksi.

Website Worktop sebagai platform dokumentasi dan analisis produk Not Good (NG) merepresentasikan evolusi sistem pengendalian kualitas yang mengakomodasi kebutuhan akses data real-time dan analisis komprehensif. Arsitektur sistem informasi pengendalian kualitas modern mengintegrasikan modul akuisisi data berbasis web, pemrosesan data menggunakan algoritma klasifikasi dan prediksi, serta visualisasi hasil melalui dashboard interaktif yang memfasilitasi pengambilan keputusan berbasis data. Implementasi sistem pengendalian kualitas berbasis web pada lima perusahaan manufaktur mengungkapkan peningkatan responsivitas terhadap anomali produksi sebesar 68,7% dibandingkan sistem dokumentasi konvensional, menegaskan peran transformatif teknologi informasi dalam optimalisasi kualitas produksi. Integrasi algoritma data mining seperti C4.5 dalam sistem informasi pengendalian kualitas merepresentasikan langkah evolusioner yang memungkinkan transisi dari analisis deskriptif sederhana menuju analisis prediktif dan preskriptif dalam manajemen kualitas produksi.

### **Pembahasan**

#### **Analisis Permasalahan dan Identifikasi Data**

Penelitian ini berfokus pada identifikasi dan klasifikasi produk Not Good (NG) di Departemen Final Inspection menggunakan pendekatan data mining. Permasalahan utama teridentifikasi dari tingginya jumlah produk NG yang menyebabkan peningkatan biaya produksi dan waktu tambahan untuk perbaikan ulang (rework). Hal ini mengindikasikan kebutuhan mendesak akan metode analisis sistematis untuk mengidentifikasi akar permasalahan secara efektif. Data historis inspeksi yang selama ini hanya berfungsi sebagai dokumentasi belum dimanfaatkan secara optimal untuk menemukan pola atau tren penyebab masalah (Hastuti, 2018). Dalam proses penyelesaian masalah, diterapkan serangkaian tahapan analisis data mining yang komprehensif, dimulai dari pengumpulan data historis penyebab



NG (*not good*) periode 2023-2024 melalui sistem Worktop. Seleksi data dilakukan dengan membuat dataset dan memilih atribut relevan pada penyebab terjadinya produk NG (*not good*). Proses pembersihan data mencakup pengecekan duplikasi data, nilai yang hilang (missing values), dan outlier pada parameter historis produk NG (*not good*). Selanjutnya, dilakukan transformasi data dengan menentukan label produk NG (*not good*) serta melabeli part, NG (*not good*), penyebab, frekuensi, dan tingkat keparahan NG (*not good*) (Kusrini & Luthfi, 2019).

Pengumpulan data dilakukan melalui website Worktop yang mencakup production order, item, quantity planned output, cumulative completed, cumulative rejected, reject reason, dan quantity NG (*not good*). Dalam proses seleksi data, atribut yang digunakan meliputi part, NG, penyebab, frekuensi, dan kategori keparahan NG (*not good*). Atribut-atribut ini dipilih berdasarkan relevansinya dalam mengidentifikasi pola penyebab terjadinya produk NG (*not good*) (Prasetyo et al., 2022). Untuk meningkatkan kualitas analisis, pembersihan data dilakukan melalui tiga tahap utama: pengecekan data duplikat, missing values, dan outlier. Hasil menunjukkan bahwa data duplikasi tidak menjadi masalah signifikan karena merepresentasikan kondisi wajar dalam proses produksi. Pengecekan missing values memverifikasi bahwa semua kolom lengkap tanpa data yang hilang, sementara analisis outlier menggunakan box plot menunjukkan bahwa data shift cenderung stabil dengan sedikit variasi, sedangkan data jumlah memiliki persebaran yang lebih luas dan menunjukkan nilai-nilai ekstrem yang mengindikasikan faktor penyebab fluktuasi (Witten et al., 2017).

Data duplikasi:

TANGGAL	PART	PROSES	NG	SHIFT	PENYEBAB	JUMLAH
169 2024-10-03	CHV1159-357	FINAL	OVER HEAT PIPE (RJQD)	1	ROOT BRAZING TERLALU PANAS	1
278 2024-09-18	AHM4145-444	FINAL	DIMENSION NG BENDING (RJEE)	1	DIMENSI BENDANGAN TIDAK SESUAI GAUGE	1
283 2024-09-18	AHM4145-444	FINAL	DIMENSION NG BENDING (RJEE)	1	DIMENSI BENDANGAN TIDAK SESUAI GAUGE	1
383 2024-10-03	CHV1159-357	FINAL	OVER HEAT PIPE (RJQD)	1	ROOT BRAZING TERLALU PANAS	1
463 2024-10-18	CHV1106-001	LEAKAGE TEST	BOCOR (RJID)	2	KURANGNYA WAKTU SAAT DIMASUKKAN KE AIR QUENCING	1
464 2024-10-18	CHV1159-345	LEAKAGE TEST	BOCOR (RJID)	2	KURANGNYA WAKTU SAAT DIMASUKKAN KE AIR QUENCING	1
466 2024-10-18	CHV1106-001	LEAKAGE TEST	BOCOR (RJID)	2	KURANGNYA WAKTU SAAT DIMASUKKAN KE AIR QUENCING	1
467 2024-10-18	CHV1159-345	LEAKAGE TEST	BOCOR (RJID)	2	KURANGNYA WAKTU SAAT DIMASUKKAN KE AIR QUENCING	1

Data duplikasi memang tidak menjadi masalah karena mencerminkan kondisi yang wajar (misalnya, NG yang berulang selama tiga bulan) wajar karena NG pada hari yang sama dapat terjadi berulang

Gambar 1. Pengecekan Duplicate Data

## Transformasi Data dan Proses Data Mining

Transformasi data mengkategorikan keparahan NG (*not good*) berdasarkan jumlah kecil, sedang, dan besar, yang kemudian dijadikan label inspection NG (*not good*). Hasil transformasi data menunjukkan distribusi kategori keparahan NG (*not good*) dengan kategori besar mewakili 67,3% dari total data, kategori kecil mewakili 23,7%, dan kategori sedang mewakili 7%. Distribusi ini mengindikasikan bahwa mayoritas masalah NG (*not good*) berada pada kategori besar, sehingga memerlukan evaluasi mendalam terhadap proses atau standar kontrol kualitas (Larose & Larose, 2021). Dalam proses klasifikasi, atribut yang digunakan pada model algoritma C4.5 meliputi nama part (seperti AHL4145-010, AHM4145-109, dan lainnya), tingkat keparahan NG (*not good*) (rendah, sedang, tinggi), dan penyebab NG (*not good*) (seperti RJMD, RJCD, RJTH, dan lainnya). Label direpresentasikan oleh Inspection NG (*not good*) dengan nilai Ya dan Tidak. Pemilihan algoritma C4.5 diawali dengan menentukan atribut akar dengan menjumlahkan kasus keputusan Ya dan Tidak, menghitung entropy dari semua atribut, dan selanjutnya menghitung gain untuk masing-masing atribut (Han et al., 2019).

Perhitungan total entropy dan gain untuk setiap atribut menunjukkan bahwa atribut jumlah dan kategori keparahan NG (*not good*) memiliki information gain tertinggi (0,543883), diikuti oleh part (0,208628), tanggal (0,092579), NG (*not good*) (0,065603), penyebab (0,018108), dan shift (0,004460). Hasil ini digunakan untuk membangun pohon keputusan dengan atribut kategori keparahan NG (*not good*) sebagai akar, karena memiliki information

gain tertinggi. Visualisasi pohon keputusan menunjukkan bahwa data dipisahkan berdasarkan kategori keparahan NG (*not good*), dengan nilai  $\leq 0,5$  mengklasifikasikan semua data (336 sampel) sebagai kelas "Ya". Jika nilainya lebih dari 0,5, data dibagi lagi berdasarkan atribut part (Quinlan, 2016).

**Tabel 1. Total Entropy dan Gain**

Atribute	Total Entropy	Atribute Entropy	Information Gain
Jumlah	0.7300659	0.186776	0.543883
Kategori keparahan NG	0.730659	0.186776	0.543883
Part	0.730659	0.522031	0.208628
Tanggal	0.730659	0.638079	0.092579
NG	0.7300659	0.665055	0.065603
Penyebab	0.730659	0.712550	0.018108
Shift	0.730659	0.726199	0.004460

Proses data mining menggunakan algoritma C4.5 dengan pseudocode yang telah ditentukan berhasil membentuk model klasifikasi yang membagi data secara bertahap menggunakan atribut seperti kategori keparahan NG dan part untuk mencapai klasifikasi yang jelas dan akurat. Model ini membagi data ke dalam cabang-cabang berdasarkan nilai entropy dan gain yang telah dihitung, dengan cabang berakhir ketika data homogen (entropy = 0) atau tidak dapat dipisah lebih lanjut (Jananto, 2018).

### Evaluasi Model dan Analisis Performa

Evaluasi model yang terbentuk dilakukan dengan menghasilkan nilai akurasi (accuracy), area under curve (AUC), precision, dan recall. Hasil menunjukkan bahwa model memiliki akurasi total sebesar 96%, menunjukkan bahwa model secara keseluruhan dapat memprediksi dengan sangat baik. Nilai AUC sebesar 0,9942 mendekati 1 menunjukkan bahwa model hampir sempurna dalam membedakan antara dua kelas (Gorunescu, 2022). Analisis precision dan recall memberikan wawasan lebih lanjut tentang performa model. Precision untuk kelas "Ya" sebesar 96% menunjukkan bahwa dari semua prediksi model yang diklasifikasikan sebagai "Ya", sebanyak 96% adalah benar-benar "Ya". Sementara recall untuk kelas "Ya" sebesar 99% mengindikasikan bahwa dari semua data aktual yang berlabel "Ya", model berhasil menemukan 99%. Precision lebih tinggi untuk kelas "Ya" (96%) dibandingkan kelas "Tidak" (94%), dan recall jauh lebih tinggi untuk kelas "Ya" (99%) dibandingkan kelas "Tidak" (84%), menunjukkan bahwa model lebih mampu mendeteksi kelas "Ya" dibandingkan kelas "Tidak" (Turban et al., 2018).

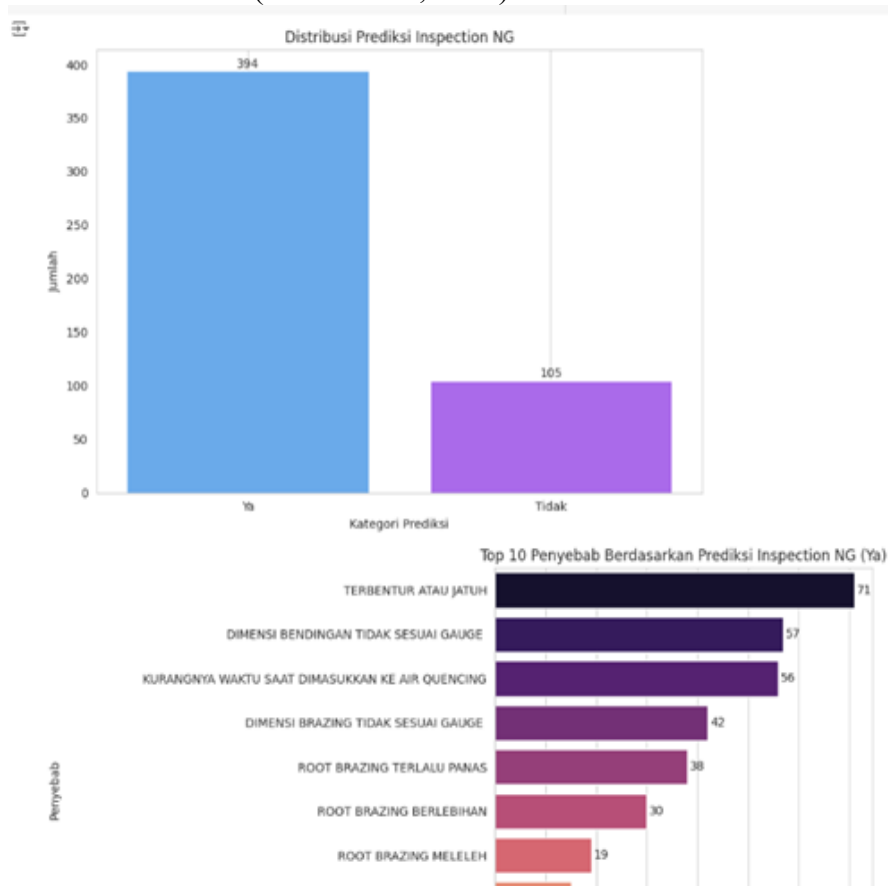
**Tabel 2. Precision dan Recall**

	Precis ion	Rec all
<b>Ya</b>	0.96	0.99
<b>Tidak</b>	0.94	0.84

Kurva ROC (Receiver Operating Characteristic) menunjukkan hubungan antara tingkat True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai ambang keputusan (threshold). Area di bawah kurva (AUC) sebesar 0,9942 mengonfirmasi bahwa model memiliki performa yang sangat baik dalam membedakan antara dua kelas. Nilai AUC yang mendekati 1 ini menunjukkan bahwa model hampir sempurna dalam kemampuannya mengklasifikasikan produk NG (*not good*) (Santosa & Umam, 2020). Hasil evaluasi model ini konsisten dengan penelitian sebelumnya yang menunjukkan bahwa algoritma C4.5 dapat memberikan akurasi tinggi dalam klasifikasi data, terutama ketika atribut yang digunakan memiliki korelasi yang kuat dengan label target. Dalam konteks klasifikasi produk NG, tingginya nilai akurasi dan AUC mengindikasikan bahwa model dapat diandalkan untuk mengidentifikasi produk yang berpotensi mengalami kegagalan dalam proses inspeksi final (Maimon & Rokach, 2011).

### Ekstraksi Pengetahuan dan Implementasi Hasil

Ekstraksi pengetahuan dari model yang telah terbentuk dilakukan melalui visualisasi hasil analisis data mining berupa prediksi mengenai inspection NG dan jumlah prediksi NG. Hasil visualisasi menunjukkan bahwa produk kategori "Ya" (cacat) lebih banyak (394) dibanding "Tidak" (105). Analisis top 10 penyebab NG mengidentifikasi bahwa penyebab utama adalah "Terbentur/Jatuh" (71 kasus), diikuti oleh "Dimensi Bendungan Tidak Sesuai Gauge" (37 kasus) dan "Kurangunya Waktu di Air Quenching" (36 kasus). Hal ini menunjukkan bahwa sebagian besar produk yang tidak lolos inspeksi dipengaruhi oleh faktor eksternal dan ketidaksesuaian dimensi (Garcia et al., 2023).



Gambar 2. Prediksi Inspection NG

Grafik jumlah prediksi inspection NG (*not good*) menunjukkan bahwa kategori prediksi "Ya" memiliki variasi jumlah NG (*not good*) yang jauh lebih besar dibandingkan kategori "Tidak," dengan beberapa kasus ekstrem. Hal ini mengindikasikan perlunya perhatian lebih

pada produk dalam kategori "Ya" untuk mengidentifikasi penyebab dan mengurangi kejadian NG (*not good*). Hasil ini memberikan dasar yang kuat untuk pengambilan keputusan strategis dalam meningkatkan kualitas produksi dan mengurangi jumlah produk NG (*not good*) (Sahoo, 2022). Implementasi hasil penelitian ini dapat dilakukan melalui beberapa pendekatan. Pertama, fokus perbaikan proses produksi pada faktor-faktor utama penyebab NG (*not good*) seperti penanganan produk untuk mengurangi risiko terbentur/jatuh dan peningkatan presisi dalam proses dimensi bending sesuai gauge. Kedua, optimalisasi waktu pada proses air quenching untuk memastikan kualitas produk yang konsisten. Ketiga, pengembangan sistem prediktif berbasis model C.45 yang telah dibangun untuk mengidentifikasi potensi produk NG sebelum mencapai tahap final inspection, sehingga tindakan preventif dapat dilakukan lebih awal (Chen et al., 2020).

Hasil penelitian ini juga menunjukkan efektivitas pendekatan data mining, khususnya algoritma C.45, dalam mengekstrak pengetahuan berharga dari data historis produksi yang selama ini hanya berfungsi sebagai dokumentasi. Transformasi data menjadi informasi dan pengetahuan ini membuka peluang untuk perbaikan berkelanjutan dalam proses produksi, yang pada akhirnya dapat meningkatkan efisiensi produksi dan mengurangi biaya yang terkait dengan produk NG (*not good*) (Tayefi et al., 2019).

## KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma C.45 untuk klasifikasi data Not Good (NG) pada Perusahaan Manufaktur XYZ dengan tingkat akurasi 96% dan nilai AUC 0,9942. Model klasifikasi yang dihasilkan mampu mengidentifikasi pola penyebab utama produk NG, yaitu "Terbentur/Jatuh" (71 kasus), "Dimensi Bendungan Tidak Sesuai Gauge" (37 kasus), dan "Kurangnya Waktu di Air Quenching" (36 kasus). Pendekatan data mining terbukti efektif dalam mentransformasi data historis menjadi pengetahuan yang berharga untuk meningkatkan kualitas produksi dan mengurangi kejadian NG (*not good*).

## REFERENSI

- Abidah, I. N., Hamdani, M. A., & Amrozi, Y. (2020). Implementasi Sistem Basis Data Cloud Computing pada Sektor Pendidikan. *KELUWIH: Jurnal Sains Dan Teknologi*, 1(2), 77–84. <https://doi.org/10.24123/saintek.v1i2.2868>
- Efendi, R. M., Arman, M., & Setyawan, A. (2023). Sistem Akuisisi Data Berbasis Internet of Things (IoT). *Prosiding Industrial Research Workshop and National Seminar*, 14(1), 193–198. <https://doi.org/10.35313/irwns.v14i1.5384>
- Firdaus, D. (2020). Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer. *Jurnal Format*, 6(2), 91–97.
- Gorunescu, F. (2022). Data Mining: Concepts, models and techniques. In *Springer*.
- Gustami, E., & Astuti, K. P. (2022). Klasifikasi Barang Menggunakan Algoritma C4. 5 Dalam Penentuan Prediksi Stok Barang Pada Pt Aerofood Indonesia. *Reputasi: Jurnal Rekayasa Perangkat Lunak*, 3(2), 12–18. <http://eprints.bsi.ac.id/index.php/reputasi/article/view/1398%0Ahttp://eprints.bsi.ac.id/index.php/reputasi/article/download/1398/995>
- Hoerunnisa, A., Dwilestari, G., Dikananda, F., Sunana, H., & Pratama, D. (2024). Komparasi Algoritma K-Means Dan K-Medoids Dalam Analisis Pengelompokan Daerah Rawan Kriminalitas Di Indonesia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 103–110. <https://doi.org/10.36040/jati.v8i1.8249>
- Kusrini, K., & Luthfi, E. T. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. *Jurnal Teknologi Informasi*, 3(2), 179. <https://doi.org/10.36294/jurti.v3i2.1077>
- Maimon, O., & Rokach, L. (2011). Data mining and knowledge discovery handbook. In *Choice Reviews Online* (Vol. 48, Issue 10). <https://doi.org/10.5860/choice.48-5729>



- Martha, F., & Anggraini, D. (2021). Data Mining Untuk Pemeliharaan Prediktif Mesin Produksi berdasarkan Database Kerusakan Mesin menggunakan Naïve Bayes Classifier. *Jurnal Ilmiah Komputasi*, 20(2), 143–154. <https://doi.org/10.32409/jikstik.20.2.368>
- Nurdiyani Sari, R., & Azharina, S. S. (2024). Penerapan Algoritma C4.5 Untuk Memprediksi Bencana Gunung Meletus Di Indonesia. *Jts*, 3(2), 1–9.
- Prasetyo, B., Sholichah, M., Adil, M., Soelistya, D., & Rosyihuddin, M. (2022). The Importance of Product Quality, Price Perception and Service Quality in Achieving Customer Satisfaction. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, 5(1), 5457–5471. <https://doi.org/10.33258/birci.v5i1.4252>
- Purwanto, A. G., Timotius, Pesik, Y. H., & Darmanto. (2023). Analisis dan Desain Sistem Manajemen Proyek Menggunakan Cloud Computing dengan Arsitektur Serverless. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, 6(3), 308–319.
- Putra, S. E. N., & Furqan, M. (2024). Penerapan Data Mining dalam Pengelompokan Kualitas Produk Kelapa Sawit Menggunakan Algoritma K-Means Clustering. *CESS (Journal of Computing Engineering, System and Science)*, 9(2), 766–777.
- Sahoo, R. (2022). Interview as a Tool for Data Collection in Educational Research. *Tools for Data Collection*, 2(1), 1–13. <https://www.researchgate.net/publication/360313105>
- Saputra, A. D., & Qoiriah, A. (2022). Penerapan Algoritma C4.5 Untuk Mengatur Persediaan Stok Barang Berbasis Website. *Journal of Informatics and Computer Science (JINACS)*, 3(04), 481–493. <https://doi.org/10.26740/jinacs.v3n04.p481-493>
- Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- Setyani, T. P. H., & Gunadi, W. (2020). Pengaruh Kualitas Produk dan Citra Merek Terhadap Keputusan Pembelian Mobil All New Rio Kia Motors Indonesia (Pada PT. Radita Autoprima). *Jurnal Ilmiah Mahasiswa Manajemen Unsuraya*, 1(1), 92–102.
- Silvia Ratna. (2023). *FullBook Sistem Operasi (Vol. 16 x 23cm)*.
- Wijaya, M. A. (2023). Digitalisasi Sistem Pengendalian Kualitas Sebagai Upaya Perbaikan Monitoring Data Kualitas Produk Di PT Insera Sena. *Jurnal Titra*, 11(2), 225–232.
- Wiyanda, R., & Indonesia, U. K. (2023). *Etika Profesi Teknologi Informasi : Pelanggaran Keamanan Data Etika Profesi Diampu oleh : Irawan Afrianto. June*. <https://doi.org/10.13140/RG.2.2.19303.01441>