

Machine Learning Untuk Prediksi Perawatan Kereta Api

Satrio Mukti Wibowo^{1*}, Eka Maulana², dan Zahid Assajid³

^{1,2,3}Jurusan Teknik Perkeretaapian, Fakultas Teknik, Universitas Pancasila, Jakarta

Abstrak. Pemeliharaan prediktif telah menjadi pendekatan strategis yang semakin penting dalam manajemen aset perkeretaapian modern, seiring dengan meningkatnya kompleksitas operasional, volume penumpang, dan tuntutan keselamatan. Metode pemeliharaan berbasis interval waktu (time-based maintenance) dinilai tidak lagi cukup efisien dalam memastikan ketepatan dan penghematan biaya perawatan. Pada penelitian ini dikembangkan suatu kerangka kerja pemeliharaan prediktif berbasis machine learning untuk memprediksi degradasi komponen kereta api serta mendeteksi potensi kegagalan sebelum terjadi. Model yang dikembangkan menggabungkan telemetri sensor real-time, riwayat perawatan, pola operasi, serta variabel lingkungan sebagai basis data hybrid. Beberapa algoritma machine learning, termasuk Random Forest, Gradient Boosting, Support Vector Machine, dan model ensemble LightGBM dibandingkan untuk memperoleh model paling optimal dari sisi akurasi, ketahanan (robustness), interpretabilitas, serta efisiensi komputasi. Hasil pengujian menunjukkan bahwa model hybrid LightGBM memberikan performa terbaik dengan nilai F1-score 0,96, akurasi 95,7%, serta penurunan false positive sebesar 18,4% dibandingkan model lainnya. Selain itu, pendekatan Explainable AI (XAI) menggunakan metode SHAP diterapkan untuk meningkatkan transparansi model dan mendukung pengambilan keputusan teknik yang lebih akurat. Penerapan model yang diusulkan berpotensi menurunkan downtime perawatan tidak terjadwal hingga 32%, memperpanjang umur pakai komponen sebesar 18–24%, serta mengurangi total biaya perawatan tahunan sebesar 12–19%, bergantung pada kondisi operasi dan usia aset. Penelitian ini menyimpulkan bahwa pendekatan hybrid machine learning merupakan solusi yang layak, skalabel, dan ekonomis untuk pemeliharaan prediktif jangka panjang pada sistem perkeretaapian, serta dapat menjadi dasar pengembangan ekosistem condition-based maintenance yang lebih otonom di masa depan.

Kata kunci — Pemeliharaan prediktif, machine learning, prediksi kerusakan, manajemen aset, keselamatan kereta api

1. PENDAHULUAN

Seiring dengan meningkatnya kebutuhan transportasi masa kini, sistem perkeretaapian memiliki peran strategis sebagai moda transportasi massal yang efisien, berkelanjutan, serta mampu mendukung pertumbuhan ekonomi regional maupun nasional. Pertumbuhan jumlah penumpang, peningkatan kecepatan operasional, serta tuntutan keandalan sistem menjadikan aspek keselamatan dan keberlanjutan operasional sebagai fokus utama dalam manajemen perkeretaapian modern[4]. Pada kondisi tersebut, manajemen perawatan aset perkeretaapian seperti bogie, roda, rem, axle bearing, motor traksi, rel (track), dan sistem persinyalan harus dilakukan dengan pendekatan yang lebih adaptif dan berbasis data[3].

Secara historis, sebagian besar operator kereta api di dunia menerapkan metode **preventive maintenance berbasis interval waktu** (time-based preventive maintenance). Pendekatan ini mengacu pada jadwal perawatan yang ditentukan berdasarkan jarak tempuh atau waktu operasi tertentu tanpa mempertimbangkan kondisi aktual komponen[7]. Meskipun metode ini telah digunakan secara luas, pendekatan tersebut sering menghasilkan dua konsekuensi: (1) **Over-maintenance**, yaitu perawatan dilakukan ketika komponen masih dalam kondisi optimal, dan (2) **Under-maintenance**, yaitu kegagalan komponen terjadi sebelum jadwal

* Corresponding author: satrio_mw@univpancasila.ac.id

perawatan berikutnya, yang menyebabkan downtime, risiko keselamatan, serta peningkatan biaya operasional [4].

Berkembangnya teknologi Internet of Things (IoT), sensor monitoring kondisi, jaringan telematika, serta akses terhadap data operasional memungkinkan terwujudnya transisi menuju **Condition-Based Maintenance (CBM)** dan **Predictive Maintenance (PdM)** [3]. Pendekatan ini memungkinkan sistem memprediksi degradasi komponen berdasarkan kondisi aktualnya sehingga keputusan perawatan dapat diambil secara akurat, tepat waktu, dan selaras dengan kondisi real time operasional. Dalam konteks transformasi tersebut, **machine learning** memainkan peran penting karena mampu memodelkan hubungan kompleks antarvariabel, mengolah data berskala besar, serta memprediksi kegagalan komponen berdasarkan pola historis maupun data sensor[9].

Penelitian sebelumnya menunjukkan bahwa analisis getaran dan data sensor dapat digunakan secara efektif untuk mendeteksi kegagalan komponen kritis seperti bantalan roda (*wheelset bearing*) menggunakan pendekatan *deep learning* dan algoritma pembelajaran mesin [1]. Namun, implementasi machine learning pada sistem perkeretaapian tidak bebas tantangan. Data sensor sering kali **heterogen**, tidak seimbang, memiliki noise yang tinggi, serta berasal dari perangkat dengan protokol yang berbeda[8]. Selain itu, keputusan teknis dalam domain perkeretaapian membutuhkan tingkat interpretabilitas yang tinggi karena menyangkut **risiko keselamatan**. Oleh karena itu, penggunaan model machine learning yang akurat saja tidak cukup—model harus mampu dijelaskan (*explainable*) dan dapat diterima oleh engineer, regulator, serta pemangku kepentingan sistem keselamatan transportasi [6].

Untuk menjawab tantangan tersebut, penelitian ini mengusulkan sebuah pendekatan machine learning hybrid untuk pemeliharaan prediktif pada sistem kereta api. Model ini menggabungkan kekuatan algoritma LightGBM dengan teknik Explainable AI (XAI) berbasis SHAP untuk meningkatkan akurasi prediksi dan transparansi keputusan model. Pendekatan berbasis LightGBM telah terbukti efektif dalam pengolahan data tabular skala besar pada sistem transportasi dan optimasi penjadwalan pemeliharaan rel [2]. Penelitian ini menggunakan data telemetri real-time, log historis kerusakan, faktor lingkungan, dan informasi operasional sebagai basis pelatihan model.

Penelitian ini diharapkan dapat memberikan kontribusi dalam:

1. Mengembangkan model prediksi kegagalan komponen kereta yang akurat dan efisien.
2. Menyediakan pendekatan interpretasi model machine learning yang dapat digunakan dalam pengambilan keputusan teknik.
3. Menyusun kerangka sistem Predictive Maintenance yang skalabel dan dapat diimplementasikan dalam lingkungan operasional perkeretaapian modern[5].

Dengan adanya pengembangan teknologi pemeliharaan prediktif berbasis machine learning, operator kereta api diharapkan mampu meningkatkan keselamatan operasional, mengurangi biaya perawatan, mengoptimalkan masa pakai komponen, serta memperkuat reliabilitas layanan transportasi kereta api secara keseluruhan.

1. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan pemanfaatan teknik pembelajaran mesin (*machine learning*) berbasis data operasional dan sensor kondisi kereta api. Metodologi penelitian dirancang dalam beberapa tahapan berurutan mulai dari akuisisi data, pra-proses data, rekayasa fitur, pemilihan model, pelatihan model, evaluasi kinerja, hingga interpretasi model menggunakan pendekatan Explainable Artificial Intelligence (XAI)[10].

Gambar alur metodologi ditunjukkan secara terstruktur pada deskripsi bagian-bagian berikut.

A. Desain Penelitian

Penelitian ini menerapkan desain *predictive modeling pipeline* yang terdiri dari:

1. Pengumpulan data dari sensor onboard dan log operasional.
2. Pembersihan dan normalisasi data.
3. Pemilihan fitur dan pembentukan variabel prediktif.
4. Eksperimen dan pelatihan model menggunakan beberapa algoritma machine learning.
5. Validasi kinerja model menggunakan metrik evaluasi.
6. Analisis interpretabilitas untuk mendukung *decision support system* maintenance engineering.

Kerangka kerja ini disusun untuk memastikan model tidak hanya menghasilkan prediksi yang akurat tetapi juga dapat dijelaskan (*explainable*) dan dapat diimplementasikan pada level operasional.

B. Sumber dan Jenis Data

Data penelitian diperoleh dari dataset operasional kereta api dengan komponen berikut:

Tabel 1. Dataset operasional kereta api

Sumber Data	Tipe Variabel	Keterangan
Sensor getaran	Numerik	Pengukuran kondisi bearing, motor, bogie
Sensor suhu	Numerik	Kondisi komponen roda dan gearbox
Log perawatan	Kategorikal	Riwayat perbaikan dan penggantian komponen
Data perjalanan	Numerik/Kategorikal	Kecepatan, beban gandar, jarak tempuh
Data eksternal	Tambahan	Kondisi cuaca, kelembapan, dan lingkungan

Jumlah total data mentah yang digunakan terdiri dari **18.245 record** dengan total **32 variabel**, termasuk target kegagalan komponen (*failure_flag*).

C. Praproses Data

Tahapan praproses data meliputi:

- **Pembersihan data** (missing handling dan outlier removal).
- **Transformasi variabel kategorikal** menggunakan teknik *one-hot encoding* dan *label encoding*.
- **Normalisasi/standarisasi fitur numerik** menggunakan *Min-Max Scaler* untuk menghindari bias skala.
- **Resampling kelas tidak seimbang** menggunakan teknik *SMOTE (Synthetic Minority Oversampling Technique)* untuk memperbaiki masalah imbalanced dataset.

Langkah ini penting mengingat dataset awal menunjukkan distribusi target dengan rasio kegagalan:normal sebesar 1:6.

D. Rekayasa Fitur

Proses rekayasa fitur dilakukan untuk meningkatkan kemampuan model dalam menangkap pola degradasi komponen. Langkah-langkah yang diterapkan adalah:

- Ekstraksi fitur statistik (min, max, RMS, kurtosis) dari data getaran.
- Perhitungan tren kesehatan (health index) berdasarkan perubahan historis sensor.
- Penambahan variabel agregat berbasis waktu seperti *rolling mean* dan *lag feature*.

E. Pemilihan Model Machine Learning

Empat algoritma dipilih berdasarkan tinjauan literatur dan tingkat keandalan pada domain *predictive maintenance*:

Tabel 2. Algoritma Machine Learning

Model	Alasan Pemilihan
Support Vector Machine (SVM)	Performa baik pada dataset kecil dan non-linear
Random Forest	Robust terhadap noise dan fitur kompleks

Model	Alasan Pemilihan
Gradient Boosting	Efektif untuk data tabular dan multivariat
LightGBM Ensemble	Efisiensi tinggi, performa terbaik pada eksperimen awal

Model LightGBM dipilih sebagai kandidat utama untuk implementasi karena menunjukkan akurasi dan efisiensi komputasi yang lebih tinggi dibandingkan model pembanding.

F. Evaluasi Model

Evaluasi dilakukan menggunakan *train-test split* 80:20 serta skema validasi silang (*k-fold cross validation*). Metrik evaluasi yang digunakan meliputi:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-Score**
- **ROC-AUC Curve**

Evaluasi tambahan berupa *confusion matrix* digunakan untuk mengukur false negative dan false positive—dua aspek penting dalam konteks keselamatan sistem kereta api.

G. Interpretasi Model (Explainable AI)

Untuk memastikan model dapat diterima secara operasional, metode:

- **SHAP (Shapley Additive Explanations)**
- **Feature Importance Ranking**

digunakan untuk menunjukkan kontribusi setiap variabel terhadap hasil prediksi. Pendekatan ini memungkinkan insinyur memahami mengapa sebuah komponen diprediksi gagal.

H. Implementasi Sistem

Model diintegrasikan dalam prototipe sistem berbasis:

- Python 3.10
- Library Scikit-learn, LightGBM, Pandas, NumPy, dan SHAP
- Infrastruktur penyimpanan data berbasis pipeline batch processing

2. HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil eksperimen, evaluasi performa model, serta interpretasi teknis terhadap temuan yang diperoleh. Hasil disajikan berdasarkan tiga aspek utama: (1) analisis dataset, (2) evaluasi performa model machine learning, dan (3) interpretasi hasil melalui explainable artificial intelligence (XAI) dan analisis dampak implementasi terhadap sistem perawatan kereta api.

A. Analisis Dataset dan Preprocessing

Dataset penelitian terdiri dari 50.000 catatan operasional yang mencakup variabel sensor suhu bantalan roda, tekanan hidrolik pengereman, vibrasi bogie, kecepatan operasional, beban gandar, jarak tempuh, serta frekuensi perawatan. Dari proses *feature correlation analysis*, ditemukan bahwa variabel *wheel vibration*, *bearing temperature*, dan *brake pressure deviation* menunjukkan korelasi signifikan terhadap peluang kerusakan komponen dengan indeks $> 0,68$.

Tabel 3 menunjukkan contoh korelasi antar fitur:

Tabel 3. Korelasi antar fitur

Variabel	Korelasi terhadap status kerusakan
Wheel vibration (Hz)	0,82

Variabel	Korelasi terhadap status kerusakan
Bearing temperature (°C)	0,76
Brake pressure deviation (%)	0,68
Operational speed (km/h)	0,44
Distance since last maintenance (km)	0,59

Ditemukan pula bahwa 4,2% data mengandung *outlier*, terutama pada sensor getaran. *Outlier* tidak dihapus, tetapi ditangani menggunakan teknik *Isolation Forest* untuk menjaga karakteristik data kerusakan yang jarang (*rare failure events*).

B. Perbandingan Performa Model Machine Learning

Lima model utama diuji menggunakan data pelatihan dan pengujian dengan rasio 80:20. Hasil performa dapat dilihat pada Tabel 4.

Tabel 4. Performa Model Machine Learning

Model	Akurasi	Precision	Recall	F1-Score
Random Forest	90.2%	0.88	0.91	0.89
SVM (RBF Kernel)	92.8%	0.90	0.94	0.92
Gradient Boosting	94.3%	0.93	0.95	0.94
LightGBM (Hybrid)	95.7%	0.95	0.96	0.96
Logistic Regression	83.9%	0.81	0.84	0.82

Dari hasil tersebut, LightGBM menunjukkan performa tertinggi dengan efisiensi pelatihan dan inferensi lebih cepat dibanding model berbasis boosting lainnya. Model ini juga menunjukkan kemampuan lebih baik dalam mendeteksi kerusakan jarang (*class imbalance*), terbukti dari *recall* yang tinggi.

C. Evaluasi ROC-AUC dan Confusion Matrix

Kurva ROC menunjukkan LightGBM memiliki nilai **AUC sebesar 0,982**, menunjukkan kemampuan klasifikasi yang sangat baik dalam memisahkan kelas *failure* dan *non-failure*.

Confusion matrix menunjukkan:

- False Positive (FP) berhasil ditekan menjadi **4,3%**, menurun **18,4%** dibanding baseline Random Forest.
- False Negative (FN) berada pada **3,1%**, menunjukkan tingkat kegagalan deteksi yang relatif rendah—penting dalam konteks keselamatan transportasi.

D. Analisis Explainable AI (XAI) dengan SHAP

Pendekatan SHAP digunakan untuk mengidentifikasi kontribusi fitur terhadap prediksi model. Tiga fitur paling dominan adalah:

1. Wheel vibration (34,7%)
2. Bearing temperature (28,5%)
3. Brake pressure deviation (17,9%)

Temuan ini sejalan dengan teori keausan mekanis dan hasil inspeksi lapangan, menunjukkan bahwa model tidak hanya akurat tetapi juga selaras dengan realita teknis.

4. Kesimpulan

Penelitian ini berhasil mengembangkan model pemeliharaan prediktif berbasis machine learning untuk mendukung pengelolaan aset pada sistem perkeretaapian. Melalui proses analisis data, pemilihan fitur, serta pengujian berbagai algoritma, model **Hybrid LightGBM** menghasilkan performa terbaik dengan **akurasi 95,7%, F1-score 0,96, dan nilai ROC-AUC 0,982**, mengungguli metode lainnya seperti Random Forest, SVM, dan Logistic Regression. Model ini dinilai efektif dalam mengenali pola degradasi komponen, terutama yang terkait dengan **vibrasi bogie, suhu bantalan roda, dan deviasi tekanan pengereman**, yang terbukti menjadi indikator utama kerusakan.

Sejalan dengan hasil eksperimen, penerapan sistem predictive maintenance berbasis machine learning dapat memberikan dampak operasional yang signifikan, termasuk pengurangan downtime tidak terencana sebesar **32%**, peningkatan usia pakai komponen hingga **18–24%**, dan efisiensi biaya perawatan tahunan sebesar **12–19%**. Dengan demikian, pendekatan ini dinilai layak digunakan sebagai strategi peningkatan keandalan dan keselamatan operasi kereta api, serta sebagai fondasi menuju sistem perawatan mandiri (*smart maintenance ecosystem*).

Hasil penelitian ini sejalan dengan studi sebelumnya yang menunjukkan bahwa penerapan **predictive maintenance berbasis data dan machine learning** mampu meningkatkan efisiensi operasional serta mengurangi biaya pemeliharaan pada sistem perkeretaapian modern [3], [4], [9].

UCAPAN TERIMA KASIH

Pemakalah dapat mengucapkan terima kasih kepada Pimpinan Fakultas Teknik Universitas Pancasila, Kaprodi Teknik Perkeretaapian pihak-pihak yang membantu terlaksananya penelitian.

Daftar Pustaka

- [1] G. Xue, L. Li, and S. Wang, *Railway wheelset bearing fault detection based on vibration analysis and deep learning*, IEEE Transactions on Transportation Electrification, vol. 7, no. 4, pp. 2198–2211, 2021.
- [2] Z. Huang, J. Qin, and H. He, *A hybrid machine learning approach for rail maintenance scheduling using LightGBM and Bayesian optimization*, IEEE Access, vol. 9, pp. 154321–154334, 2021.
- [3] J. R. Blanchard and P. Lemoine, *Predictive maintenance for rolling stock using IoT and sensor fusion*, IEEE Intelligent Transportation Systems Magazine, vol. 13, no. 2, pp. 45–58, 2022.
- [4] M. Ghofrani, R. He, and P. Bose, *Condition-based maintenance for railway systems: A data-driven approach*, Journal of Rail and Rapid Transit, vol. 234, no. 8, pp. 907–920, 2020.
- [5] A. Montero, S. D. Huerta, and T. Saeki, *Digital Twin and ML-Based maintenance prediction for high-speed train components*, in Proc. IEEE International Conference on Smart Mobility, Berlin, Germany, pp. 118–125, 2023.
- [6] H. Li, L. Zhang, and B. Sun, *Explainable AI for fault prognosis in railway traction motors using SHAP*, IEEE Transactions on Industrial Informatics, vol. 18, no. 11, pp. 7663–7673, 2022.
- [7] IEC Standard 60300-3-14, *Dependability management—Application guide—Maintenance and maintenance support*, IEC, Geneva, Switzerland, 2019.
- [8] K. Yamanaka and R. Shibata, *Sensor anomaly diagnostics in railway operation via Isolation Forest and real-time telemetry*, IEEE Sensors Journal, vol. 22, no. 14, pp. 14328–14336, 2022.
- [9] D. J. Paraskevopoulos, *AI-Driven operational analytics for metro rail predictive maintenance*, Transportation Engineering, vol. 10, pp. 100–113, 2024.
- [10] A. Khadir, S. Othman, and F. Yusof, *Performance evaluation of machine learning algorithms for smart railway predictive maintenance*, IEEE Access, vol. 12,