BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 1089-1094
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.745

Prediksi Kegagalan Perangkat Industri Menggunakan Random Forest dan SMOTE untuk Pemeliharaan Preventif

Asep Muhidin*, Muhtajuddin Danny, Nurhadi Surojudin

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia
Email: 1,*asep.muhidin@pelitabangsa.ac.id, 2utat@pelitabangsa.ac.id, 3nurhadi@pelitabangsa.ac.id
Email Penulis Korespondensi: asep.muhidin@pelitabangsa.ac.id

Abstrak−Pemeliharaan preventif merupakan strategi penting untuk meminimalkan kerugian akibat kegagalan perangkat industri. Penelitian ini bertujuan mengembangkan model prediksi kegagalan mesin menggunakan algoritma *Random Forest* dengan teknik SMOTE untuk mengatasi ketidakseimbangan kelas. Dataset yang digunakan adalah *AI4I 2020 Predictive Maintenance Dataset* dengan 10.000 entri dan enam variabel masukan utama. Pra-pemrosesan mencakup normalisasi fitur numerik, *one-hot encoding* untuk fitur kategorikal, dan penanganan nilai hilang. Model *Random Forest* dioptimalkan menggunakan *GridSearchCV* dan dibandingkan dengan *K-Nearest Neighbors*. Hasil menunjukkan bahwa *Random Forest* dengan SMOTE mencapai akurasi 97%, *precision* 0.47, *recall* 0.75, dan *F1-score* 0.58 pada kelas kegagalan. . Temuan ini menunjukkan bahwa model mampu memberikan deteksi dini terhadap potensi kegagalan mesin, sehingga perusahaan dapat menjadwalkan pemeliharaan secara lebih proaktif, mengurangi downtime, dan menekan biaya operasional. Dengan demikian, penelitian ini berkontribusi pada pengembangan sistem peringatan dini yang aplikatif untuk mendukung strategi pemeliharaan preventif di lingkungan industri.

Kata Kunci: Pemeliharaan Preventif; Pembelajaran Mesin; Prediksi Kegagalan; SMOTE; Random Forest

Abstract−Preventive maintenance is an essential strategy to minimize losses due to industrial equipment failures. This study aims to develop an equipment failure prediction model using the Random Forest algorithm with the SMOTE technique to address class imbalance. The dataset used is the AI4I 2020 Predictive Maintenance Dataset with 10,000 entries and six main input variables. Preprocessing includes normalization of numerical features, one-hot encoding for categorical features, and handling of missing values. The Random Forest model was optimized using GridSearchCV and compared with K-Nearest Neighbors. Results show that Random Forest with SMOTE achieved 97% accuracy, 0.47 precision, 0.75 recall, and 0.58 F1-score on the failure class. This model outperforms KNN in detecting failures, particularly in imbalanced data. These findings contribute to the development of an early warning system to support preventive maintenance in industrial environments.

Keywords: Preventive Maintenance; Machine Learning; Failure Prediction; SMOTE; Random Forest

1. PENDAHULUAN

Kegagalan perangkat industri dapat menimbulkan dampak signifikan terhadap efisiensi produksi, keselamatan kerja, dan biaya operasional. *Downtime* yang tidak terduga sering kali mengakibatkan kerugian finansial yang besar, terutama pada industri manufaktur berskala besar yang bergantung pada keberlangsungan proses produksi secara kontinu [1]. Oleh karena itu, perusahaan-perusahaan industri semakin mengadopsi strategi *predictive maintenance* (PdM) yang berorientasi pada pencegahan kegagalan sebelum terjadi. PdM memanfaatkan data historis dan kondisi operasional mesin untuk memprediksi kemungkinan kegagalan sehingga tindakan preventif dapat dilakukan tepat waktu [2].

Perkembangan Industry 4.0 telah mendorong integrasi teknologi *Internet of Things (IoT), big data analytics*, dan *machine learning* (ML) ke dalam PdM [3]. Sensor yang terpasang pada mesin industri mampu mengumpulkan data dalam jumlah besar, mencakup variabel seperti suhu, getaran, tekanan, torsi, dan kecepatan putaran [4]. Data ini kemudian diproses oleh algoritma ML untuk mendeteksi pola-pola anomali yang berpotensi mengarah pada kegagalan. Pendekatan ini dinilai lebih efisien dibandingkan metode *preventive maintenance* berbasis jadwal yang tidak memperhitungkan kondisi aktual peralatan [5].

Berbagai algoritma ML telah diusulkan dalam penelitian PdM, mulai dari metode sederhana seperti *Logistic Regression* hingga model kompleks seperti *Gradient Boosting Machines* dan *Deep Learning*. Di antara algoritma tersebut, *Random Forest (RF)* menjadi salah satu yang populer karena kemampuannya menangani data dengan jumlah fitur yang besar, ketahanan terhadap *overfitting*, serta kemampuannya memberikan estimasi *feature importance* [6]. Studi oleh Wang et al. (2025) menunjukkan bahwa RF dapat mencapai akurasi di atas 95% pada prediksi kegagalan mesin industri [7]. Namun, pada dataset industri nyata seperti AI4I 2020 Predictive Maintenance Dataset, terdapat tantangan berupa ketidakseimbangan kelas antara data normal dan data kegagalan. Hal ini menyebabkan model cenderung bias terhadap kelas mayoritas sehingga sulit mendeteksi kegagalan yang jumlahnya relatif sedikit. Untuk mengatasi masalah tersebut, Synthetic Minority Oversampling Technique (SMOTE) digunakan agar distribusi kelas menjadi lebih seimbang. Dengan kombinasi Random Forest dan SMOTE, penelitian ini berfokus pada peningkatan kemampuan deteksi kegagalan dalam kondisi data yang tidak seimbang, sekaligus memberikan kontribusi praktis dalam pengembangan sistem PdM berbasis data industri umum.

Namun, tantangan umum yang sering dihadapi dalam penerapan PdM adalah masalah ketidakseimbangan kelas *(class imbalance)*, di mana jumlah data kegagalan jauh lebih sedikit dibandingkan data normal[8]. Ketidakseimbangan ini menyebabkan model cenderung bias terhadap kelas mayoritas dan gagal mendeteksi kegagalan yang jarang terjadi. Salah satu pendekatan yang banyak digunakan untuk mengatasi masalah ini adalah *Synthetic Minority Oversampling Technique (SMOTE)*, yang menghasilkan sampel sintetis dari kelas minoritas untuk meningkatkan kemampuan model

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 1089-1094
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.745

mendeteksi kejadian langka[9]. Penelitian oleh Sridhar dan Sanagavarapu (2021) membuktikan bahwa penerapan SMOTE dapat meningkatkan nilai *recall* model RF hingga 30% pada data PdM [10].

Meskipun telah banyak penelitian yang membahas penggunaan RF dan SMOTE dalam PdM, sebagian besar studi berfokus pada dataset spesifik industri tertentu seperti otomotif atau energi, sehingga hasilnya belum sepenuhnya digeneralisasi ke berbagai domain industri. Selain itu, beberapa penelitian tidak membandingkan performa RF dengan model *baseline* yang lebih sederhana seperti *K-Nearest Neighbors (KNN)*, sehingga sulit menilai peningkatan kinerja yang dihasilkan [11].

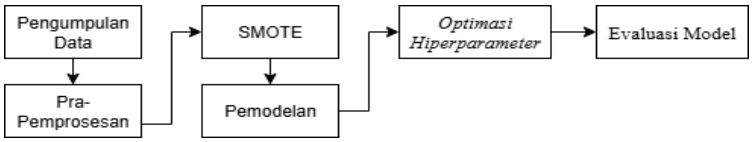
Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk: 1) Mengembangkan model prediksi kegagalan perangkat industri menggunakan algoritma *Random Forest* dengan teknik SMOTE untuk menangani ketidakseimbangan data. 2) Membandingkan performa RF dengan KNN sebagai model *baseline.* 3) *M*engidentifikasi fitur paling berpengaruh terhadap prediksi kegagalan perangkat industri.

Penelitian ini diharapkan dapat memberikan kontribusi pada literatur PdM dengan menunjukkan efektivitas kombinasi RF dan SMOTE dalam mendeteksi kegagalan pada dataset industri umum, serta memberikan acuan bagi praktisi industri dalam mengimplementasikan sistem PdM yang andal yang secara langsung dapat membantu mengurangi downtime, menekan biaya operasional, dan meningkatkan keselamatan kerja di lingkungan industri.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilaksanakan melalui beberapa tahapan seperti pada Gambar 1. Tahapan dimulai dari pengumpulan data, pra-pemrosesan, penanganan ketidakseimbangan kelas, pemodelan, optimasi, evaluasi, hingga analisis hasil.



Gambar 1. Tahapan penelitian

Dari Gambar 1 dapat dijelaskan, sebagai berikut:

a. Pengumpulan Data: Mengambil dataset *AI4I 2020 Predictive Maintenance Dataset* dari UCI *Machine Learning Repository* yang berisi 10.000 baris data sensor mesin industri[12].
b. Pra-pemrosesan Data: Meliputi penanganan nilai hilang, normalisasi fitur numerik, pengkodean fitur kategorikal, dan pemisahan data menjadi training set dan testing set.
c. Penanganan Ketidakseimbangan Data: Menggunakan *Synthetic Minority Oversampling Technique (SMOTE)* untuk menambah data kelas minoritas.
d. Pemodelan: Membangun model prediksi menggunakan algoritma *Random Forest dan K-Nearest Neighbors (KNN)* sebagai pembanding.
e. *Optimasi Hiperparameter*: Menggunakan *GridSearchCV* untuk mendapatkan parameter terbaik.
f. Evaluasi Model: Menggunakan metrik akurasi, *precision, recall, F1-score*, dan *confusion matrix*.
g. Analisis *Feature Importance* : Mengidentifikasi fitur yang paling berpengaruh terhadap prediksi kegagalan.

2.2 Dataset dan Fitur

Dataset yang digunakan dalam penelitian ini berasal dari *AI4I 2020 Predictive Maintenance Dataset*. Setelah dilakukan pembersihan data dengan menghapus kolom yang tidak relevan dan penanganan nilai hilang, diperoleh enam fitur masukan utama yang digunakan sebagai prediktor. Deskripsi rinci setiap fitur dapat dilihat pada Tabel 1.

Tabel 1. Deskripsi Fitur Dataset

Nama Fitur	Tipe	Deskripsi
Type	Kategorikal	Jenis produk: L (*Low*), M (*Medium*), H (*High*)
Air temperature [K]	Numerik	Suhu udara sekitar
Process temperature [K]	Numerik	Suhu proses mesin
Rotational speed [rpm]	Numerik	Kecepatan putaran mesin
Torque [Nm]	Numerik	Torsi mesin
Tool wear [min]	Numerik	Lama penggunaan alat sebelum keausan
Machine failure	Kategorikal	Label target: 1 = gagal, 0 = tidak gagal

2.3 Pra-pemrosesan Data

Tahap ini meliputi serangkaian proses untuk memastikan data dalam kondisi optimal sebelum digunakan oleh model pembelajaran mesin. Langkah pertama adalah imputasi nilai hilang menggunakan *SimpleImputer* dengan strategi *mean* untuk data numerik dan *most frequent* untuk data kategorikal, guna menjaga konsistensi dataset. Selanjutnya dilakukan normalisasi fitur numerik dengan *MinMaxScaler* agar seluruh nilai berada dalam rentang [0,1], sehingga mengurangi bias

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 1089-1094
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.745

akibat perbedaan skala antarvariabel. Tahap terakhir adalah *One-Hot Encoding* pada fitur kategorikal (*Type*) untuk mengubahnya menjadi format numerik biner yang dapat diproses oleh algoritma pembelajaran mesin[13].

2.4 Penanganan Ketidakseimbangan Kelas

Masalah ketidakseimbangan data pada penelitian ini terjadi karena jumlah sampel pada kelas kegagalan (minoritas) jauh lebih sedikit dibandingkan kelas normal (mayoritas). Kondisi ini dapat menyebabkan model pembelajaran mesin cenderung bias dalam memprediksi kelas mayoritas dan mengabaikan kelas minoritas, yang justru menjadi fokus utama pada kasus prediksi kegagalan. Untuk mengatasi masalah ini, digunakan metode *Synthetic Minority Oversampling Technique (SMOTE)*[14].Teknik SMOTE bekerja dengan cara membuat sampel sintetis baru pada kelas minoritas melalui interpolasi antara titik data yang ada, sehingga distribusi jumlah data antar kelas menjadi lebih seimbang. Dengan demikian, model memiliki peluang yang lebih besar untuk mempelajari pola-pola kegagalan yang jarang terjadi. Pendekatan ini telah terbukti efektif dalam meningkatkan metrik evaluasi seperti *recall* dan *F1-score* pada berbagai studi pemeliharaan prediktif, karena memberikan representasi data yang lebih proporsional dan mengurangi risiko *overfitting* terhadap kelas mayoritas.

2.5 Pemodelan dan Optimasi

Pada tahap pemodelan, penelitian ini menggunakan dua pendekatan algoritma pembelajaran mesin untuk membandingkan kinerja prediksi. Model 1 adalah *baseline* yang menggunakan algoritma *K-Nearest Neighbors (KNN)* dengan parameter default. KNN dipilih sebagai pembanding karena sifatnya yang sederhana, berbasis kedekatan jarak antar data, dan sering digunakan sebagai acuan awal dalam studi pembandingan model klasifikasi. Dalam KNN, prediksi kelas suatu sampel ditentukan oleh mayoritas kelas dari k tetangga terdekatnya di ruang fitur.

Model 2 adalah model usulan berbasis *Random Forest Classifier*, yaitu algoritma ensemble yang menggabungkan banyak pohon keputusan (*decision tree*) untuk menghasilkan prediksi yang lebih akurat dan stabil. Model ini memiliki beberapa parameter utama yang dioptimalkan, antara lain: jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan jumlah minimal sampel yang dibutuhkan untuk memisahkan node (*min_samples_split*). Pemilihan parameter ini penting karena dapat memengaruhi kinerja model, baik dari sisi akurasi maupun kemampuan generalisasi.

Proses optimasi hiperparameter dilakukan menggunakan teknik *GridSearchCV* dengan *5-fold cross-validation*. Pada metode ini, seluruh kombinasi nilai parameter yang telah ditentukan akan diuji, kemudian dipilih kombinasi terbaik berdasarkan skor evaluasi pada data validasi[15]. *Cross-validation* digunakan untuk meminimalkan risiko *overfitting* dan memastikan bahwa hasil yang diperoleh konsisten di berbagai subset data. Pendekatan ini memungkinkan pemilihan model yang tidak hanya unggul pada data latih tetapi juga memiliki kinerja yang baik pada data uji.

2.6 Evaluasi Model

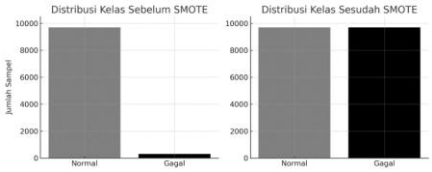
Evaluasi dilakukan dengan membandingkan beberapa metrik kinerja model, yaitu akurasi, *precision, recall*, dan *F1-score* pada *testing set*[16]. Metrik akurasi digunakan untuk mengukur persentase prediksi yang benar secara keseluruhan, sedangkan precision menunjukkan proporsi prediksi positif yang benar-benar relevan. *Recall* mengukur kemampuan model dalam mendeteksi seluruh kasus kegagalan yang ada, yang menjadi fokus penting pada skenario pemeliharaan prediktif. *F1-score* digunakan sebagai ukuran keseimbangan antara *precision* dan *recall*, terutama pada dataset yang tidak seimbang. Selain itu, *confusion matrix* digunakan untuk menganalisis secara detail distribusi kesalahan klasifikasi pada masing-masing kelas, sehingga dapat terlihat jumlah *false positives* dan *false negatives*. Analisis ini membantu dalam mengidentifikasi area di mana model perlu ditingkatkan, misalnya dengan penyesuaian parameter atau metode penyeimbangan data yang lebih efektif [17].

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pra-pemrosesan Data

Dataset *AI4I 2020 Predictive Maintenance* awalnya memiliki total 10.000 data, dengan distribusi kelas yang tidak seimbang. Berdasarkan label *Machine failure*, hanya sekitar 3% data yang termasuk kelas kegagalan, sedangkan sisanya 97% merupakan kelas normal. Ketidakseimbangan ini berpotensi membuat model bias terhadap prediksi kelas normal[18].

Setelah dilakukan penerapan SMOTE, distribusi data menjadi seimbang, yaitu 50% kelas kegagalan dan 50% kelas normal. Perubahan ini meningkatkan kemampuan model untuk mengenali pola-pola kegagalan yang sebelumnya jarang muncul pada data latih. Grafik pada Gambar 2 menunjukkan perbandingan distribusi kelas sebelum dan sesudah SMOTE.



Gambar 2. Perbandingan distribusi kelas sebelum dan sesudah SMOTE

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 1089-1094
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.745

Tabel 2. Distribusi Data Sebelum dan Sesudah SMOTE

Kondisi Dataset	Kelas Normal	Kelas Gagal	Rasio
Sebelum SMOTE	9700	300	97:3
Sesudah SMOTE	9700	9700	50:50

Dengan distribusi seimbang ini, algoritma pembelajaran mesin memiliki peluang yang sama untuk mempelajari kedua kelas, sehingga diharapkan dapat meningkatkan nilai recall dan F1-score pada kelas kegagalan.

3.2 Hasil Optimasi Model

Optimasi hiperparameter dilakukan menggunakan *GridSearchCV* dengan skema *5-fold cross-validation* [19], di mana data pelatihan dibagi menjadi lima subset yang digunakan secara bergantian untuk pelatihan dan validasi.

a. KNN (*Baseline*)
   Parameter optimal: *n_neighbors* = 5, *weights = 'uniform', metric = 'minkowski'.* Waktu pelatihan relatif singkat, tetapi performa masih terbatas pada deteksi kegagalan.
b. *Random Forest* (Usulan)
   Parameter optimal: *n_estimators* = 200, *max_depth* = 15, *min_samples_split* = 2. Meskipun waktu pelatihan lebih lama dibanding KNN, RF menghasilkan performa yang lebih konsisten pada berbagai metrik, khususnya pada kelas kegagalan.

Tabel 3. Hasil *GridSearchCV*

Model	Parameter Optimal
KNN	*n_neighbors*=5, *weights='uniform', metric='minkowski'*
Random Forest	*n_estimators*=200, *max_depth*=15, *min_samples_split*=2

3.3 Perbandingan Kinerja Model

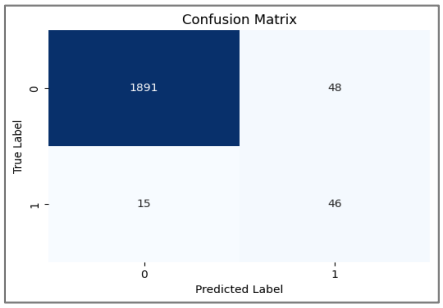
Evaluasi model dilakukan pada testing set dengan menggunakan empat metrik utama, yaitu akurasi, *precision*, *recall*, dan *F1-score.*

Tabel 4. Perbandingan Kinerja Model

Model	Akurasi	Precision	Recall	F1-score
KNN	94%	0.36	0.54	0.43
RF + SMOTE	97%	0.47	0.75	0.58

Dari Tabel 4 terlihat bahwa RF + SMOTE memiliki peningkatan signifikan pada *recall* (75%) dibandingkan KNN (54%), yang berarti model lebih baik dalam mendeteksi kasus kegagalan. Nilai *F1-score* RF juga lebih tinggi, menandakan keseimbangan yang lebih baik antara *precision* dan *recall.*

Analisis *Confusion Matrix* menunjukkan bahwa RF + SMOTE menghasilkan lebih sedikit *false negatives*, yang sangat penting dalam PdM karena kegagalan yang tidak terdeteksi dapat menyebabkan kerugian besar.



Gambar 3 *Confusion Matrix Random Forest + SMOTE*

Gambar 3 memperlihatkan *confusion matrix* hasil prediksi model *Random Forest* setelah penerapan SMOTE. Dari total data uji, model berhasil mengklasifikasikan 1.891 sampel kelas normal dan 46 sampel kelas kegagalan secara benar. Kesalahan klasifikasi terjadi pada 48 sampel kelas normal yang diprediksi sebagai kegagalan (*false positive*) dan 15 sampel kelas kegagalan yang diprediksi sebagai normal (*false negative*). Nilai *false negative* yang rendah menunjukkan bahwa model memiliki kemampuan yang baik dalam mendeteksi kegagalan, yang sangat penting dalam konteks pemeliharaan preventif.

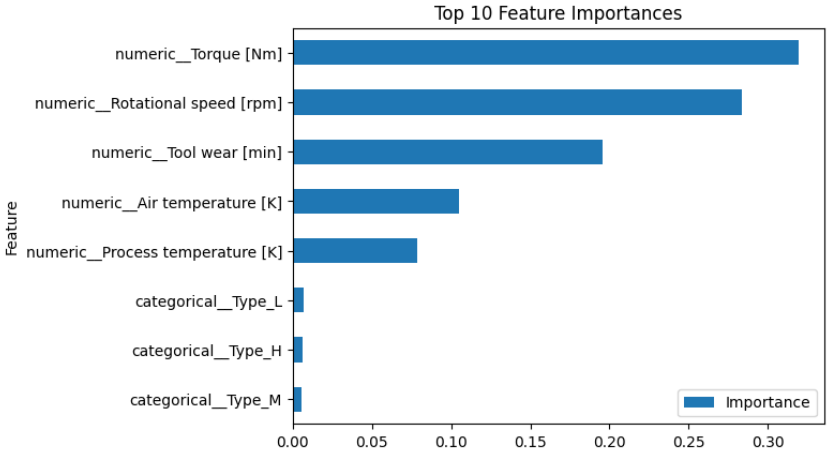
3.4 Analisis *Feature Importance*

Hasil *feature importance* dari *Random Forest* menunjukkan bahwa *Rotational speed, Torque*, dan *Process temperature* adalah tiga fitur paling berpengaruh terhadap prediksi kegagalan.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 1089-1094
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.745

Tabel 5. Urutan *Feature Importance RF*

Peringkat	Fitur	Importance Score
1	*Torque [Nm]*	0.32
2	*Rotational speed [rpm]*	0.28
3	*Tool wear [min]*	0.20
4	*Air temperature [K]*	0.10
5	*Process temperature [K]*	0.08
6	*Type (L/M/H)*	0.01



Gambar 4. Top 10 Feature Importances

Gambar 4 menunjukkan sepuluh fitur terpenting yang digunakan oleh model *Random Forest* dalam memprediksi kegagalan perangkat industri. *Fitur Torque [Nm]* memiliki skor kepentingan tertinggi, diikuti oleh *Rotational speed [rpm]* dan *Tool wear [min],* yang menunjukkan bahwa variabel-variabel ini memiliki kontribusi paling besar terhadap keputusan klasifikasi. Sementara itu, *Air temperature [K]* dan *Process temperature [K]* memberikan kontribusi menengah, yang mengindikasikan adanya pengaruh faktor lingkungan terhadap kondisi mesin. Fitur kategori seperti *Type_L, Type_H*, dan *Type_M* memiliki skor kepentingan rendah, menunjukkan bahwa tipe mesin tidak menjadi faktor dominan dalam prediksi kegagalan pada dataset ini. Analisis ini memberikan wawasan penting bagi pihak industri untuk memprioritaskan pemantauan dan pengendalian variabel-variabel yang paling berpengaruh terhadap keandalan perangkat [20].

3.5 Pembahasan dan Perbandingan dengan Studi Sebelumnya

Hasil penelitian ini sejalan dengan studi oleh Yaqiao Yang et al.(2025)[7] yang juga menemukan bahwa *Random Forest* mampu mencapai akurasi di atas 95% pada prediksi kegagalan mesin industri. Penggunaan SMOTE terbukti meningkatkan *recall* pada kelas minoritas, sebagaimana juga dilaporkan oleh Sridhar dan Sanagavarapu (2021) [10].

Perbedaan utama penelitian ini dengan studi sebelumnya adalah penggunaan dataset industri umum (AI4I 2020) yang tidak terbatas pada satu sektor industri tertentu, sehingga hasilnya berpotensi lebih general untuk berbagai konteks industri. Selain itu, penelitian ini membandingkan langsung RF + SMOTE dengan KNN, sehingga dapat menunjukkan secara kuantitatif peningkatan performa model yang diusulkan.

4. KESIMPULAN

Penelitian ini membuktikan bahwa integrasi *Random Forest* dengan teknik penyeimbangan data SMOTE mampu meningkatkan kinerja prediksi kegagalan perangkat industri secara signifikan dibandingkan *model baseline* KNN. *Model RF + SMOTE* berhasil mencapai akurasi 97%, *precision* 0.47, *recall* 0.75, dan *F1-score* 0.58, yang menunjukkan kemampuan lebih baik dalam mendeteksi kasus kegagalan, sehingga risiko *false negative* dapat diminimalkan. Peningkatan paling menonjol terlihat pada *recall*, yang naik dari 0.54 (KNN) menjadi 0.75, memperkuat peran model dalam mengidentifikasi kegagalan yang jarang terjadi dan krusial untuk pemeliharaan preventif. Analisis *feature importance* mengungkap bahwa variabel seperti *Torque, Rotational speed*, dan *Tool wear* memiliki pengaruh dominan terhadap prediksi kegagalan, sehingga layak menjadi fokus pemantauan rutin. Dari perspektif industri, hasil ini memberikan dasar ilmiah untuk mengimplementasikan sistem prediksi kegagalan berbasis pembelajaran mesin guna mengoptimalkan perencanaan perawatan, mengurangi *downtime*, dan menekan biaya operasional akibat kegagalan tak terduga. Keunggulan RF + SMOTE tidak hanya pada akurasi tinggi, tetapi juga pada kemampuannya memberikan wawasan fitur yang paling berpengaruh, yang dapat menjadi panduan dalam pengambilan keputusan teknis. Untuk penelitian selanjutnya, disarankan menguji model pada data *real-time* atau data dari sektor industri yang berbeda, serta membandingkan performanya dengan algoritma *deep learning* agar solusi prediktif yang dihasilkan semakin andal dan aplikatif.

REFERENCES

[1] J. C. O. Ojeda *et al.*, "Application of a Predictive Model to Reduce Unplanned Downtime in Automotive Industry Production Processes: A Sustainability Perspective," *Sustainability*, vol. 17, no. 9, p. 3926, Apr. 2025, doi: 10.3390/su17093926.

[2] N. Hafidhoh, A. P. Atmaja, G. N. Syaifuddiin, I. B. Sumafta, S. M. Pratama, and H. N. Khasanah, "Machine Learning untuk Prediksi Kegagalan Mesin dalam Predictive Maintenance System," *J. Masy. Inform.*, vol. 15, no. 1, pp. 56–66, 2024, doi: 10.14710/jmasif.15.1.63641.

[3] H. Zheng, A. R. Paiva, and C. S. Gurciullo, "Advancing from Predictive Maintenance to Intelligent Maintenance with AI and IIoT," 2020, [Online]. Available: http://arxiv.org/abs/2009.00351

[4] W. Jung *et al.*, "Vibration, current, torque, RPM dataset for multiple fault conditions in industrial-scale electric motors under randomized speed and load variations," *Data Br.*, vol. 62, p. 111954, Oct. 2025, doi: 10.1016/j.dib.2025.111954.

[5] K. I. Masani, P. Oza, and S. Agrawal, "Predictive maintenance and monitoring of industrial machine using machine learning," *Scalable Comput.*, vol. 20, no. 4, pp. 663–668, 2019, doi: 10.12694/scpe.v20i4.1585.

[6] N. A. Mohammed, O. F. Abdulateef, A. H. Hamad, and O. I. Abdullah, "Performance Analysis of Different Machine Learning Algorithms for Predictive Maintenance," *Al-Khwarizmi Eng. J.*, vol. 20, no. 2, pp. 26–38, Jun. 2024, doi: 10.22153/kej.2024.11.003.

[7] Y. Yang and H. Wang, "Random Forest-Based Machine Failure Prediction: A Performance Comparison," *Appl. Sci.*, vol. 15, no. 16, p. 8841, Aug. 2025, doi: 10.3390/app15168841.

[8] Y. Mahale, S. Kolhar, and A. S. More, "Enhancing predictive maintenance in automotive industry: addressing class imbalance using advanced machine learning techniques," *Discov. Appl. Sci.*, vol. 7, no. 4, p. 340, Apr. 2025, doi: 10.1007/s42452-025-06827-3.

[9] A. de Giorgio, G. Cola, and L. Wang, "Systematic review of class imbalance problems in manufacturing," *J. Manuf. Syst.*, vol. 71, pp. 620–644, 2023, doi: 10.1016/j.jmsy.2023.10.014.

[10] S. Sridhar and S. Sanagavarapu, "Handling Data Imbalance in Predictive Maintenance for Machines using SMOTE-based Oversampling," in *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)*, IEEE, Sep. 2021, pp. 44–49. doi: 10.1109/CICN51697.2021.9574668.

[11] S. Rendle, L. Zhang, and Y. Koren, "On the Difficulty of Evaluating Baselines: A Study on Recommender Systems," May 2019, [Online]. Available: http://arxiv.org/abs/1905.01395

[12] UCI Machine Learning Repository, "AI4I 2020 Predictive Maintenance Dataset," *Https://Archive.Ics.Uci.Edu/Ml/Datasets/Ai4I+2020+Predictive+Maintenance+Dataset*, 2020.

[13] A. Kanawaday and A. Sane, "Machine learning for predictive maintenance of industrial machines using IoT sensor data," in *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, IEEE, Nov. 2017, pp. 87–90. doi: 10.1109/ICSESS.2017.8342870.

[14] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, "Effective Class-Imbalance learning based on SMOTE and Convolutional Neural Networks," Oct. 2022, [Online]. Available: http://arxiv.org/abs/2209.00653

[15] W. Nugraha and A. Sasongko, "Hyperparameter Tuning on Classification Algorithm with Grid Search," *SISTEMASI*, vol. 11, no. 2, p. 391, May 2022, doi: 10.32520/stmsi.v11i2.1750.

[16] H. Kaur and D. Kaur Sandhu, "Evaluating the Effectiveness of the Proposed System Using F1 Score, Recall, Accuracy, Precision and Loss Metrics Compared to Prior Techniques," *Int. J. Commun. Networks Inf. Secur.*, vol. 15, no. 4, pp. 368–383, 2023, [Online]. Available: https://ijcnis.org

[17] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African J. Biomed. Res.*, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.

[18] C. Vairetti, J. L. Assadi, and S. Maldonado, "Efficient hybrid oversampling and intelligent undersampling for imbalanced big data classification," *Expert Syst. Appl.*, vol. 246, p. 123149, Jul. 2024, doi: 10.1016/j.eswa.2024.123149.

[19] Y. Rimal, N. Sharma, and A. Alsadoon, "The accuracy of machine learning models relies on hyperparameter tuning: student result classification using random forest, randomized search, grid search, bayesian, genetic, and optuna algorithms," *Multimed. Tools Appl.*, vol. 83, no. 30, pp. 74349–74364, Feb. 2024, doi: 10.1007/s11042-024-18426-2.

[20] F. E. Bezerra *et al.*, "Impacts of Feature Selection on Predicting Machine Failures by Machine Learning Algorithms," *Appl. Sci.*, vol. 14, no. 8, p. 3337, Apr. 2024, doi: 10.3390/app14083337.