

IMPLEMENTASI ANALISIS PREDIKTIF UNTUK DETEKSI DINI KERUSAKAN *BEARING* PADA *AIR COMPRESSOR*

Umi Yuliatin^{1*}, Aris Muamar²

¹Teknik Instrumentasi Kilang, Politeknik Energi dan Mineral Akamigas, Cepu

²PT. Kilang Pertamina Internasional, Cilacap

*E-mail: umi.yuliatin@esdm.go.id

ABSTRAK

Kerusakan bearing pada kompresor udara dapat menyebabkan *downtime* yang mahal dalam industri menyebabkan waktu maintenance yang diperlukan untuk perbaikan. Penelitian ini bertujuan untuk mengembangkan model analisis prediktif untuk deteksi dini kerusakan bearing pada kompresor udara menggunakan dua metode: Regresi Logistik dan *Random Forest*. Data sensor yang digunakan mencakup berbagai parameter operasional seperti suhu, torsi, daya motor, kecepatan rotasi, getaran, dan kebisingan. Hasil evaluasi model menunjukkan akurasi 76.5% untuk kedua metode, dengan *Random Forest* menunjukkan kinerja yang lebih baik dalam mendeteksi kondisi *Noisy* (kerusakan bearing), dengan Precision (1) sebesar 78.3% dan Recall (1) sebesar 96.1%. Sementara itu, kemampuan kedua model dalam mendeteksi kondisi *Ok* (bearing sehat) cenderung lebih rendah. Analisis distribusi dan korelasi fitur mengidentifikasi getaran dan suhu outlet sebagai parameter yang paling relevan dalam membedakan kondisi bearing *Ok* dan *Noisy*. Penelitian ini menyarankan penggunaan *Random Forest* dan Regresi logistik untuk sistem *predictive maintenance*, yang dapat membantu meningkatkan efisiensi perawatan dan mengurangi *downtime air compressor*. Temuan ini juga membuka peluang untuk pengembangan lebih lanjut menggunakan *deep learning* dan metode lain guna meningkatkan akurasi deteksi kerusakan pada komponen kritis mesin.

Kata kunci: Analisis Prediktif, Regresi Logistik, *Air Compressor*, *Random Forest*, *Predictive Maintenance*

1. PENDAHULUAN

Dalam industri modern, efisiensi dan keandalan sistem mesin sangat bergantung pada kondisi operasional yang optimal. Salah satu komponen penting yang sering digunakan dalam berbagai jenis mesin adalah kompresor udara. Kompresor udara berperan krusial dalam banyak aplikasi industri, mulai dari proses penggerak hingga sistem pendingin. Kerusakan pada kompresor udara, terutama pada komponen bearing-nya, dapat mengarah pada *downtime* yang mahal dan kerugian ekonomi yang signifikan [1].

Deteksi dini kerusakan bearing menjadi semakin penting seiring dengan kompleksitas mesin yang semakin meningkat. Salah satu metode yang menjanjikan untuk mendeteksi kerusakan adalah melalui analisis prediktif yang memanfaatkan teknik pembelajaran mesin dan data sensor. Penelitian sebelumnya menunjukkan bahwa penggunaan jaringan saraf dalam, seperti *Long Short-Term Memory* (LSTM), untuk analisis vibrasi, dapat mencapai tingkat akurasi yang tinggi dalam mendeteksi anomali pada bearing [2]. Selain itu, kombinasi sinyal vibrasi dan arus mampu memberikan wawasan yang lebih mendalam tentang kondisi mesin [3].

Berbagai jenis sensor, seperti sensor vibrasi, arus, dan suhu, telah terbukti efektif dalam pemantauan kondisi motor induksi dan komponen terkait [4]. Penggunaan teknik pengolahan sinyal, seperti analisis gelombang dan penggabungan data, memungkinkan evaluasi yang lebih

komprehensif dari tekanan dan performa kompresor [5]. Dengan demikian, penerapan analisis prediktif tidak hanya meningkatkan tingkat deteksi kerusakan tetapi juga memungkinkan perawatan yang lebih tepat waktu, yang pada gilirannya meningkatkan efisiensi operasional [6] [7].

Kinerja kompresor udara sangat ditentukan oleh interaksi kompleks antara beban mekanik, efisiensi termal, dan dinamika sistem yang stabil. Ketidaksesuaian antara beban puntir yang diberikan dan daya motor yang dihasilkan, khususnya ketika tidak sejalan dengan kecepatan rotasi kompresor, dapat mengindikasikan adanya peningkatan gesekan internal serta keausan pada elemen putar. Fluktuasi dalam kecepatan operasi sering kali disertai dengan lonjakan suhu yang disebabkan oleh pelumasan dan pendinginan yang tidak optimal [8]. Keadaan termal yang tidak konsisten, seperti kenaikan suhu oli atau keluaran udara, menandakan akumulasi panas yang berlebihan, yang dapat mempercepat kerusakan pada komponen mekanik [4]. Oleh karena itu, sistem pendingin yang tidak efisien lebih cenderung menyebabkan deteriorasi pada kompresor, mengingat ketidakmampuannya untuk menjaga keseimbangan termal yang dibutuhkan dalam operasi mesin [6].

Lebih dari itu, pola getaran dan tingkat kebisingan berperan sebagai indikator penting dalam mendeteksi gangguan struktural dan kerusakan pada bearing pada tahap awal. Peningkatan amplitudo getaran yang terdeteksi pada sumbu tertentu sering kali menunjukkan adanya ketidakseimbangan rotasi atau cacat pada permukaan bearing [9]. Kombinasi dari faktor-faktor mekanik, termal, dan dinamis ini menyediakan gambaran menyeluruh mengenai keadaan operasional kompresor udara. Adanya interaksi multidimensional antarparameter tersebut membuka peluang bagi pengaplikasian analisis prediktif untuk mendeteksi pola anomali secara dini, mendukung strategi pemeliharaan yang lebih efisien dan berbasis data [1].

Penelitian ini bertujuan untuk: (1) membandingkan kinerja Regresi Logistik dan *Random Forest* dalam mendeteksi kerusakan bearing pada kompresor udara, (2) mengidentifikasi parameter sensor yang paling berpengaruh melalui analisis *feature importance*, dan (3) memberikan rekomendasi model yang optimal untuk sistem *predictive maintenance*.

2. METODE

Dalam analisis prediktif ini digunakan dua model machine learning yang sangat populer dalam sistem klasifikasi. Kedua mode ini menunjukkan perhitungan prediksi klasifikasi dari data histori yang ada apakah terjadi *Noisy* atau tidak pada sistem bearing di mesin *air compressor*. Diharapkan kedua metode ini mampu menjawab kebutuhan dengan memberikan angka evaluasi prediksi yang tinggi.

A. Logistic Regression

Regresi logistik merupakan salah satu metode statistik klasik yang digunakan untuk memprediksi probabilitas terjadinya suatu peristiwa biner berdasarkan satu atau lebih variabel independen. Metode ini secara luas diterapkan dalam bidang *predictive maintenance* untuk mengidentifikasi kemungkinan kegagalan komponen berdasarkan parameter sensor dan kondisi operasi [10]. Tidak seperti regresi linear yang mengasumsikan hubungan linier langsung antara variabel bebas dan hasil keluaran kontinu, regresi logistik mengubah nilai keluaran menjadi probabilitas yang dibatasi antara 0 dan 1 melalui fungsi logistik atau fungsi sigmoid.

Secara matematis, hubungan antara probabilitas kejadian $P(Y=1)$ dan prediktor X_i dinyatakan sebagai:

$$P(Y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (1)$$

Persamaan tersebut dapat ditulis ulang dalam bentuk logit (*log odds*):

$$\ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (2)$$

di mana β_i adalah koefisien regresi yang menunjukkan arah dan kekuatan pengaruh setiap variabel independen terhadap peluang terjadinya suatu kejadian. Nilai positif dari β_i menunjukkan bahwa peningkatan variabel X_i meningkatkan kemungkinan kejadian (misalnya kondisi *faulty*), sedangkan nilai negatif menunjukkan hubungan yang berlawanan [10].

Dalam konteks deteksi dini kerusakan kompresor, regresi logistik dapat digunakan untuk mengestimasi peluang kerusakan bearing berdasarkan kombinasi parameter mekanik, termal, dan getaran. Sebagai contoh, peningkatan daya motor dan suhu oli yang diikuti oleh kenaikan amplitudo getaran vertikal dapat meningkatkan nilai peluang kerusakan $P(Y=1)$. Dengan demikian, model ini tidak hanya berfungsi sebagai alat prediksi, tetapi juga sebagai instrumen diagnosis yang menjelaskan faktor penyebab (*causal factors*) yang berkontribusi terhadap degradasi sistem.

Keunggulan utama regresi logistik dalam penelitian ini terletak pada interpretabilitas dan transparansinya. Setiap koefisien model dapat dikonversi menjadi rasio peluang (*odds ratio*), yang menunjukkan seberapa besar peningkatan peluang kerusakan ketika suatu variabel meningkat satu satuan. Pendekatan ini memberikan pemahaman yang lebih mudah bagi teknisi dan manajer perawatan dalam mengambil keputusan berbasis bukti, seperti menentukan ambang batas suhu, daya, atau getaran yang aman. Selain itu, regresi logistik mampu berfungsi sebagai model dasar pembandingan terhadap algoritma *machine learning* yang lebih kompleks seperti Decision Tree dan Random Forest, sehingga dapat mengevaluasi apakah hubungan linier sudah cukup menjelaskan fenomena atau diperlukan model nonlinier yang lebih adaptif [11].

Dengan demikian, regresi logistik berperan penting dalam tahap awal analisis prediktif pada sistem industri karena memberikan keseimbangan antara akurasi dan interpretabilitas. Dalam konteks *predictive maintenance* air compressor, metode ini mampu menjelaskan secara kuantitatif bagaimana perubahan nilai sensor tertentu meningkatkan risiko kegagalan bearing, sekaligus menjadi dasar perbandingan yang objektif terhadap model berbasis pohon keputusan dan ansambel.

B. Random Forest

Random Forest merupakan algoritma *ensemble learning* yang dikembangkan sebagai pengembangan dari metode *Classification and Regression Tree (CART)*. Algoritma ini menggabungkan sejumlah pohon keputusan (*Decision Trees*) yang dibangun secara acak dari subset data dan variabel untuk menghasilkan prediksi yang lebih stabil, akurat, dan tahan terhadap *overfitting*. Berbeda dengan *Decision Tree* tunggal yang sensitif terhadap variasi data pelatihan, *Random Forest* memanfaatkan prinsip *bagging (bootstrap aggregating)*, yaitu pengambilan sampel secara acak dengan pengembalian untuk membentuk kumpulan model independen. Hasil akhir klasifikasi diperoleh melalui proses *majority voting*, sedangkan untuk regresi digunakan rata-rata nilai prediksi dari seluruh pohon. Secara matematis, jika terdapat (B) pohon keputusan, maka hasil prediksi akhir dari Random Forest untuk suatu sampel (x) dirumuskan sebagai:

$$\hat{f}_{RF}(x) = \frac{1}{B} \sum b = \frac{1}{B} \sum \hat{f}_b(x) \quad (3)$$

di mana $\hat{f}_b(x)$ adalah hasil prediksi dari pohon ke- b . Untuk klasifikasi biner seperti deteksi kondisi *bearing normal* dan *faulty*, hasil akhir ditentukan berdasarkan kelas dengan frekuensi tertinggi di antara seluruh pohon [12].

Selama proses pelatihan, Random Forest juga menerapkan mekanisme random feature selection, yaitu hanya sebagian kecil dari seluruh fitur yang dipilih secara acak untuk digunakan pada setiap percabangan pohon. Strategi ini mengurangi korelasi antar pohon sehingga menghasilkan model yang lebih general dan tidak bias terhadap fitur dominan [13]. Penggunaan subset acak baik pada tingkat data maupun fitur menjadikan Random Forest mampu menangkap hubungan nonlinier yang kompleks antarparameter tanpa kehilangan kemampuan generalisasi. Salah satu keunggulan utama *Random Forest* dibanding *Decision Tree* adalah kemampuannya menyediakan feature importance, yaitu ukuran kontribusi setiap variabel terhadap akurasi model. Nilai pentingnya fitur dihitung berdasarkan penurunan rata-rata *Gini impurity* atau *Information Gain* yang dihasilkan oleh fitur tersebut di seluruh pohon dalam model. Fitur dengan nilai *importance* tinggi dianggap memiliki pengaruh signifikan terhadap keluaran model. Dalam konteks sistem kompresor udara, analisis ini dapat digunakan untuk menentukan parameter sensor yang paling berkontribusi terhadap kerusakan bearing, seperti kombinasi beban mekanik, suhu pelumasan, dan amplitudo getaran.

Dalam implementasi *predictive maintenance*, Random Forest memberikan keseimbangan ideal antara akurasi tinggi dan kemampuan interpretasi teknis. Berbagai penelitian terdahulu menunjukkan bahwa model ini efektif dalam mengidentifikasi pola degradasi komponen dan mendeteksi anomali multivariat pada sistem rotasi [11]. Dalam penelitian ini penulis menggunakan Decission tree dan random forest untuk klasifikasi bearing ari compressor, serta menilai sejauh mana model berbasis ansambel mampu meningkatkan performa deteksi dini kerusakan bearing melalui analisis multi-sensor.

3. PEMBAHASAN

Dataset yang digunakan dalam penelitian ini merupakan dataset sekunder sistem kompresor udara yang diambil dari platform Kaggle. Dataset ini terdiri dari 1000 entri data yang mencakup 26 kolom, dengan rincian 21 kolom berjenis data numerik dan 5 kolom lainnya berjenis data kategorikal. Setiap entri menggambarkan kondisi operasional kompresor udara dalam siklus tertentu, yang mencakup pengukuran berbagai parameter fisik seperti torsi, daya motor, kecepatan rotasi, suhu, getaran, dan kebisingan.

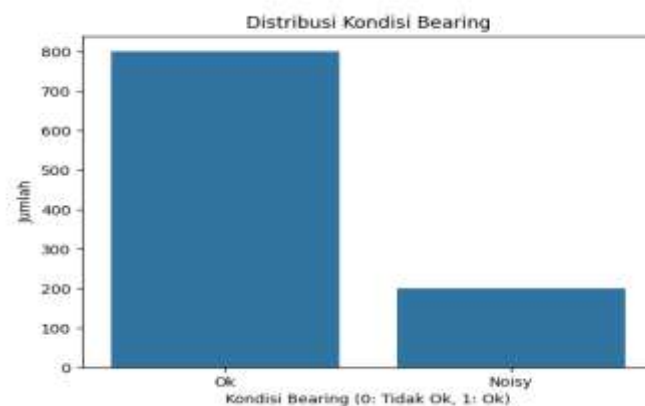
Data numerik dalam dataset ini mencakup variabel-variabel seperti suhu, daya motor, tekanan udara keluaran, getaran pada sumbu-sumbu tertentu, dan banyak lagi. Sementara itu, data kategorikal mencakup status kondisi komponen seperti bearing, pompa, radiator, dan exhaust valve, yang menunjukkan apakah komponen tersebut dalam kondisi baik (Ok) atau rusak (*Faulty*). Tabel 1 berikut adalah ringkasan statistika data set yang digunakan untuk keperluan analisis :

Tabel 1. Parameter yang digunakan dalam analisis.

Jenis Data	Parameter	Mean	Min	Max	Deskripsi
Numerik	RPM	1,499.52 RPM	480 RPM	2,520 RPM	Variasi besar dalam kecepatan rotasi motor.
Numerik	Torque	49.19 Nm	13.22 Nm	93.54 Nm	Rentang besar pada torsi motor, menunjukkan kondisi beban yang bervariasi.
Numerik	Noise (dB)	53.41 dB	39.88dB	93.53 dB	Tingkat kebisingan kondisi mekanis

Numerik	Outlet Temp	118.85°C	76.90°C	160°C	Menunjukkan variasi suhu di outlet, berhubungan dengan kondisi operasi sistem.
Kategorikal	Bearing	-	Ok	Tidak Ok	Variabel target,
Kategorikal	AC Motor	Stable	Stable	Stable	Semua nilai menunjukkan kondisi stabil, tidak relevan untuk analisis lebih lanjut.

Gambar 1 menunjukkan distribusi kondisi bearing pada sistem kompresor udara, dengan sebagian besar data (sekitar 800 entri) menunjukkan kondisi *Ok*, menandakan bahwa kompresor beroperasi dalam keadaan normal. Namun, sekitar 200 entri menunjukkan kondisi *Noisy*, yang mengindikasikan adanya gangguan atau kerusakan pada bearing, yang sering kali terkait dengan peningkatan getaran dan penurunan efisiensi. Ketidakseimbangan distribusi ini menunjukkan pentingnya model analisis prediktif untuk mendeteksi kerusakan dini pada bearing yang *Noisy*, guna mengoptimalkan perawatan dan mengurangi downtime pada sistem kompresor.

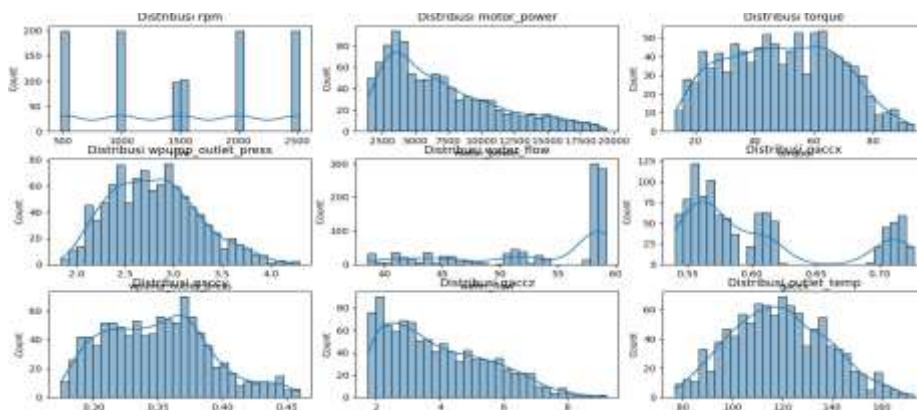


Gambar 1. Jumlah Distribusi Data Bearing

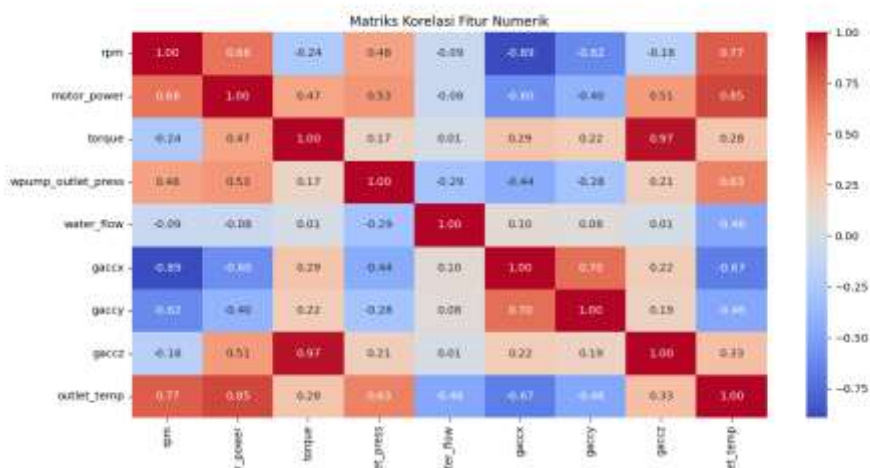
Gambar 2 menunjukkan distribusi histogram dari berbagai parameter operasional kompresor udara, yang memberikan gambaran tentang variasi dan karakteristik data sensor. Beberapa parameter seperti RPM, motor power, dan outlet temperature menunjukkan distribusi yang lebih normal, dengan puncak di nilai tengah, yang mengindikasikan kestabilan dan konsistensi dalam kinerja sistem kompresor. Sebaliknya, parameter seperti *wpump_outlet_press* dan *water_flow* menunjukkan distribusi yang lebih terdistribusi dengan puncak pada nilai ekstrem, mencerminkan variasi yang lebih besar dalam kondisi operasi yang tidak selalu stabil. Getaran pada sumbu X dan Z (*gaccx*, *gaccz*) serta torque menunjukkan distribusi dengan beberapa puncak yang lebih tajam, mengindikasikan adanya fluktuasi beban atau gangguan yang mempengaruhi kestabilan kompresor. Secara keseluruhan, distribusi ini menggambarkan keragaman dalam kondisi operasi yang dapat digunakan untuk lebih memahami hubungan antar parameter dalam memprediksi kegagalan komponen, terutama bearing.

Gambar 3 menunjukkan hubungan antar berbagai parameter numerik dalam sistem kompresor udara. Secara umum, RPM memiliki korelasi positif yang kuat dengan motor power (0.85) dan outlet temperature (0.77), menunjukkan bahwa semakin tinggi kecepatan rotasi motor, semakin tinggi daya motor dan suhu keluaran udara kompresor. Torque memiliki korelasi tinggi dengan motor power (0.47) dan outlet temperature (0.28), yang mengindikasikan bahwa torsi motor yang lebih tinggi berhubungan dengan peningkatan daya dan suhu. Sementara itu, parameter getaran (*gaccx*, *gaccy*, *gaccz*) menunjukkan korelasi negatif dengan parameter lain seperti RPM dan motor power, terutama untuk *gaccx* yang berkorelasi negatif

kuat dengan RPM (-0.89), menandakan bahwa peningkatan getaran pada sumbu X berbanding terbalik dengan kecepatan rotasi motor. *Water flow* menunjukkan korelasi rendah dengan sebagian besar parameter lainnya, menunjukkan bahwa aliran air tidak banyak dipengaruhi oleh beban atau kondisi operasi lainnya. Analisis korelasi ini memberikan gambaran tentang hubungan fisis antarparameter dan bisa digunakan untuk memahami faktor-faktor yang mempengaruhi kinerja sistem kompresor secara lebih holistik.



Gambar 2. Distribusi Data Variabel Terkait

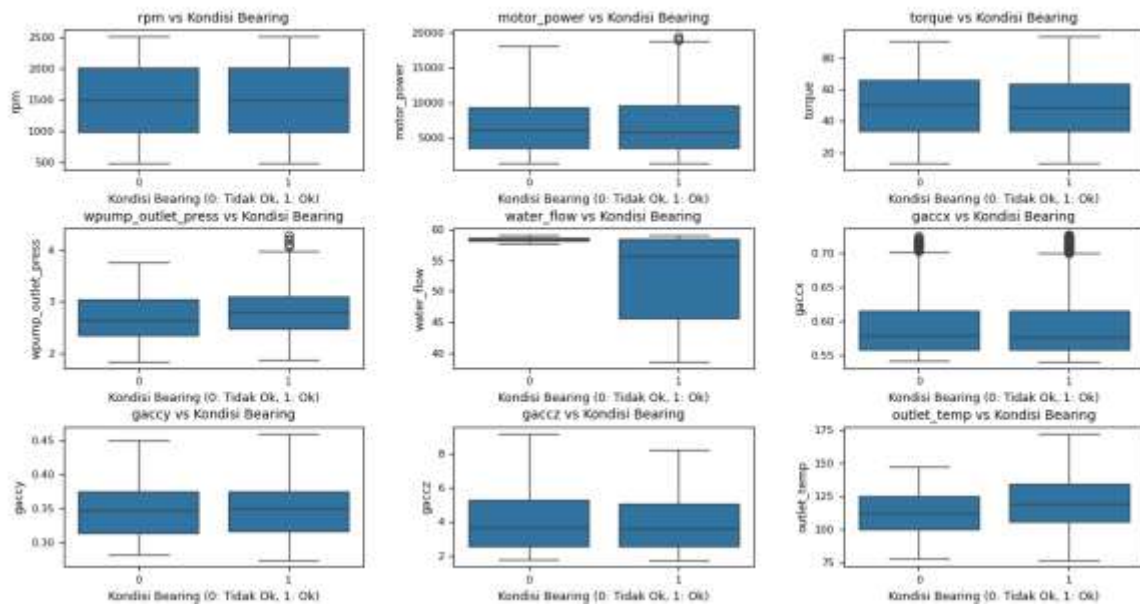


Gambar 3. Heatmap Matrik Korelasi

Gambar 4 menunjukkan hubungan antara fitur operasional kompresor dan kondisi bearing, membandingkan distribusi nilai tiap parameter untuk dua kategori target: *Ok* dan *Noisy*. Parameter seperti RPM, motor power, dan torque menunjukkan distribusi serupa antara kedua kondisi bearing, tanpa perbedaan yang signifikan. Sementara itu, water flow dan wpump_outlet_press menunjukkan perbedaan lebih jelas, dengan kondisi bearing *Noisy* memiliki nilai yang lebih tinggi, yang mungkin mengindikasikan gangguan pada sistem aliran atau tekanan yang mempengaruhi kondisi bearing.

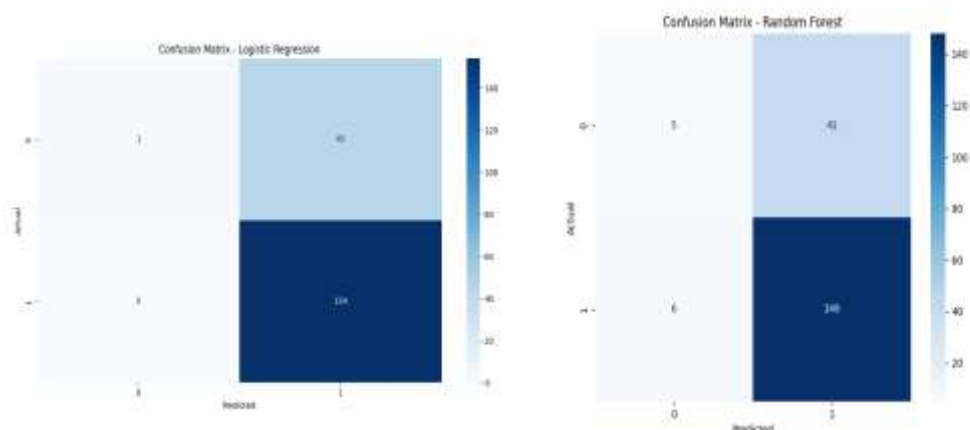
Selain itu, parameter getaran (gaccx, gaccy, gaccz) menunjukkan variasi yang lebih tinggi pada kondisi bearing *Noisy*, khususnya pada sumbu gaccx. Suhu outlet (outlet_temp) juga sedikit lebih tinggi pada kondisi *Noisy*, yang mencerminkan hubungan antara suhu berlebih dan kerusakan bearing. Secara keseluruhan, getaran dan suhu outlet terlihat sebagai indikator yang

lebih signifikan dalam membedakan kondisi bearing yang sehat dan rusak, sedangkan parameter lain seperti torsi dan daya motor menunjukkan hubungan yang lebih lemah.



Gambar 4. Hubungan Fitur dengan Target

Hasil confusion matrix pada Gambar 5 menunjukkan bahwa pada model *Logistic Regression*, sebagian besar sampel dengan kondisi bearing *Ok* berhasil diidentifikasi dengan benar sebagai *Ok* (True Positive = 154) dan tidak terdapat sampel *Ok* yang diklasifikasikan sebagai *Noisy* (False Negative = 0). Namun demikian, kemampuan model dalam mengenali kondisi bermasalah masih sangat rendah, terlihat dari hanya satu sampel *Noisy* yang berhasil diprediksi sebagai *Noisy* (True Negative = 1), sementara 45 sampel *Noisy* lainnya justru diklasifikasikan sebagai *Ok* (False Positive = 45). Pola ini mengindikasikan bahwa model *Logistic Regression* cenderung bias terhadap kelas *Ok* dan kurang sensitif dalam mendeteksi adanya *noise* pada bearing.



Gambar 5. Confusion Matrix dari Klasifikasi Regresi Logistik dan *Random Forest*

Pada model Random Forest, kecenderungan serupa masih tampak, meskipun dengan perbaikan tertentu. Sebagian besar sampel bearing *Ok* tetap diprediksi dengan benar sebagai *Ok* (True Positive = 148), namun mulai muncul sejumlah sampel *Ok* yang diklasifikasikan sebagai *Noisy* (False Negative = 6), yang mencerminkan adanya alarm palsu terhadap kondisi yang sebenarnya masih normal. Di sisi lain, kemampuan model dalam mengenali kondisi *Noisy* sedikit meningkat dibanding Logistic Regression, yang ditunjukkan oleh bertambahnya jumlah True Negative menjadi 5 sampel, meskipun 41 sampel *Noisy* masih keliru diprediksi sebagai *Ok* (False Positive = 41).

Secara keseluruhan, kedua model menunjukkan performa yang relatif baik dalam mengklasifikasikan kondisi bearing yang normal, tetapi masih memiliki keterbatasan signifikan dalam mendeteksi kondisi *Noisy*. Dominannya nilai False Positive pada kedua model mengindikasikan bahwa banyak kasus bearing bermasalah justru terklasifikasi sebagai *Ok*, sehingga berpotensi menimbulkan risiko terlewatnya indikasi gangguan pada sistem. Temuan ini menegaskan perlunya optimasi lebih lanjut, baik melalui penyeimbangan kelas, pemilihan fitur yang lebih representatif, maupun penyesuaian ambang keputusan (*decision threshold*), agar model lebih peka terhadap gejala *noise* pada bearing dan lebih andal digunakan analisis ini. Tabel 2 berikut kriteria evaluasi model secara keseluruhan :

Tabel 2 . Evaluasi Model

Model	Akurasi	<i>Precision</i> (1)	<i>Recall</i> (1)	<i>F1-Score</i> (1)	<i>Precision</i> (0)	<i>Recall</i> (0)	<i>F1-Score</i> (0)
<i>Random Forest</i>	0.765	0.783	0.961	0.863	0.454	0.109	0.175
<i>Logistic Regression</i>	0.775	0.773	1	0.873	1	0.217	0.042

Berdasarkan hasil evaluasi yang disajikan dalam tabel, kedua model menunjukkan nilai akurasi yang relatif berdekatan, yaitu 0,775 untuk Logistic Regression dan 0,765 untuk Random Forest. Namun, jika ditinjau berdasarkan metrik per kelas, tampak bahwa kinerja keduanya lebih unggul pada kelas 1 (kondisi bearing *Ok*) dibandingkan kelas 0 (kondisi bearing *Noisy*). Untuk kelas 1, baik Logistic Regression maupun Random Forest memiliki nilai precision, recall, dan F1-Score yang tinggi, yang menunjukkan kemampuan yang sangat baik dalam mengenali kondisi bearing yang normal. Logistic Regression sedikit lebih unggul dari sisi recall karena mampu mengklasifikasikan hampir seluruh sampel *Ok* dengan benar, sedangkan Random Forest menunjukkan kombinasi precision dan recall yang juga cukup seimbang untuk kelas tersebut.

Secara keseluruhan, tabel tersebut menunjukkan bahwa model cenderung bias terhadap kelas *Ok* (kelas 1): kinerja sangat baik untuk mendeteksi kondisi normal, tetapi sensitivitas terhadap kelas *Noisy* masih sangat terbatas. Hal ini tercermin dari nilai precision, recall, dan F1-Score yang jauh lebih rendah pada kelas 0, baik untuk Random Forest maupun Regresi Logistik, yang mengindikasikan bahwa banyak kondisi bearing *Noisy* yang tidak terdeteksi secara memadai. Dalam konteks deteksi dini gangguan pada bearing, temuan ini menegaskan perlunya strategi tambahan, seperti penyeimbangan kelas, pemilihan fitur yang lebih representatif, atau penyesuaian ambang keputusan (*decision threshold*), agar model lebih peka terhadap kondisi *Noisy* yang secara praktis justru lebih kritis untuk mendukung implementasi pemeliharaan prediktif yang andal.

Perbedaan akurasi antara Random Forest (76,5%) dan Logistic Regression (77,5%) secara absolut hanya sebesar 1 poin persentase, atau setara dengan selisih dua sampel benar dari total 200 data uji menunjukkan perbedaan akurasi tersebut tidak dapat dianggap signifikan secara statistik; secara praktis, performa global kedua model dapat dikatakan relatif setara. Hal ini bisa disebabkan oleh ketidakseimbangan kelas yang menyebabkan rendahnya sensitivitas terhadap

kelas Noisy, sehingga disarankan penggunaan teknik *Synthetic Minority Over-sampling Technique (SMOTE)* untuk menyeimbangkan distribusi kelas dan menguji sejauh mana peningkatan representasi kelas Noisy dapat memperbaiki metrik evaluasi, khususnya recall dan F1-Score kelas *Noisy*.

4. SIMPULAN

Penelitian ini berhasil mengembangkan dua model prediktif, yaitu Regresi Logistik dan Random Forest, untuk mendeteksi kondisi bearing pada kompresor udara berbasis data sensor. Kedua model menunjukkan akurasi yang relatif berdekatan, yakni 77,5% untuk Regresi Logistik dan 76,5% untuk Random Forest; selisih sekitar satu poin persentase (setara dua sampel benar dari total 200 data uji) secara praktis tidak dapat dianggap signifikan secara statistik, sehingga kinerja global keduanya dapat dikatakan sebanding. Analisis metrik per kelas menunjukkan bahwa kedua model lebih unggul dalam mengenali kondisi bearing *Ok* (kelas 1), dengan nilai precision, recall, dan F1-Score yang tinggi, sementara sensitivitas terhadap kondisi *Noisy* (kelas 0) masih terbatas yang tercermin dari nilai recall dan F1-Score yang rendah pada kelas tersebut. Temuan ini mengindikasikan bahwa model cenderung bias terhadap kelas *Ok* dan masih sering gagal mendeteksi kondisi bearing *Noisy*, sehingga dalam konteks deteksi dini gangguan, aspek yang perlu lebih diperhatikan bukan hanya perbandingan akurasi, tetapi juga kemampuan model dalam mengidentifikasi kelas bermasalah. Kondisi ini diduga kuat berkaitan dengan ketidakseimbangan kelas pada data, sehingga pada pengembangan selanjutnya disarankan penerapan teknik *Synthetic Minority Over-sampling Technique (SMOTE)*, penyeimbangan kelas, penyesuaian ambang keputusan (decision threshold), serta optimasi fitur untuk meningkatkan sensitivitas dan F1-Score pada kelas *Noisy*.

Analisis distribusi dan korelasi fitur juga menunjukkan bahwa parameter getaran dan suhu outlet merupakan variabel yang paling berpengaruh dalam membedakan kondisi bearing *Ok* dan *Noisy*, sehingga keduanya berpotensi dijadikan indikator utama dalam sistem *predictive maintenance*. Implikasi praktis dari hasil ini adalah tersedianya dasar pemodelan awal untuk meningkatkan efisiensi perawatan kompresor udara melalui deteksi dini indikasi kerusakan bearing, dengan memanfaatkan kombinasi informasi getaran dan suhu outlet sebagai sinyal peringatan. Meskipun demikian, penelitian ini masih memiliki ruang pengembangan melalui penggunaan data yang lebih besar dan beragam, penerapan teknik penanganan ketidakseimbangan data seperti SMOTE secara sistematis, serta eksplorasi metode machine learning lainnya agar diperoleh model dengan akurasi, sensitivitas terhadap kondisi *Noisy*, dan keandalan prediksi yang lebih tinggi dalam lingkungan operasional nyata.

5. DAFTAR PUSTAKA

- [1] A. Ciobotaru, C. Corches, D.-I. Gota, and L. Miclea, "An Explainable Deep Learning-Based Predictive Maintenance Solution for Air Compressor Condition Monitoring," *Sensors*, 2025, doi: 10.3390/s25185797.
- [2] B. G. Joung, C. Nath, Z. Li, and J. W. Sutherland, "Bearing Anomaly Detection in an Air Compressor Using an LSTM and RNN-based Machine Learning Model," *Int. J. Adv. Manuf. Technol.*, 2024, doi: 10.1007/s00170-024-14322-z.
- [3] S. Ayankoso, A. Dutta, Y. He, F. Gu, A. D. Ball, and S. K. Pal, "Performance of Vibration and Current Signals in the Fault Diagnosis of Induction Motors Using Deep Learning and Machine Learning Techniques," *Struct. Heal. Monit.*, 2024, doi: 10.1177/14759217241289874.
- [4] O. Mey *et al.*, "Condition Monitoring of Drive Trains by Data Fusion of Acoustic Emission and Vibration Sensors," *Processes*, 2021, doi: 10.3390/pr9071108.
- [5] J. Liu, Y. Hu, Y. Wang, B. Wu, J. Fan, and Z. Hu, "An Integrated Multi-Sensor Fusion-Based Deep

- Feature Learning Approach for Rotating Machinery Diagnosis,” *Meas. Sci. Technol.*, 2018, doi: 10.1088/1361-6501/aaaca6.
- [6] P. Dwivedi *et al.*, “Predictive Maintenance and Monitoring of Industrial Compressors Using Machine Learning: A Proactive Approach,” *Metall. Mater. Eng.*, 2025, doi: 10.63278/1387.
- [7] J. Mohammed and J. Abdulhady, “Rolling Bearing Fault Detection Based on Vibration Signal Analysis and Cumulative Sum Control Chart,” *Fme Trans.*, 2021, doi: 10.5937/fme2103684m.
- [8] C. Ates, T. Höfchen, M. Witt, R. Koch, and H. J. Bauer, “Vibration-Based Wear Condition Estimation of Journal Bearings Using Convolutional Autoencoders,” *Sensors*, vol. 23, no. 22, 2023, doi: 10.3390/s23229212.
- [9] S. R. Saufi, Z. A. Bin Ahmad, M. S. Leong, and M. H. Lim, “Low-Speed Bearing Fault Diagnosis Based on ArSSAE Model Using Acoustic Emission and Vibration Signals,” *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2909756.
- [10] A. J. Scott, D. W. Hosmer, and S. Lemeshow, “Applied Logistic Regression,” *Biometrics*, vol. 47, no. 4, 1991, doi: 10.2307/2532419.
- [11] N. Hadroug, A. Hafaifa, A. Iratni, and M. Guemana, “Reliability Modeling Using an Adaptive Neuro-Fuzzy Inference System: Gas Turbine Application,” *Fuzzy Inf. Eng.*, 2021, doi: 10.1080/16168658.2021.1915451.
- [12] G. Biau and E. Scornet, “A random forest guided tour,” *Test*, vol. 25, no. 2, 2016, doi: 10.1007/s11749-016-0481-7.
- [13] V. Svetnik, A. Liaw, C. Tong, J. Christopher Culberson, R. P. Sheridan, and B. P. Feuston, “Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling,” *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 6, 2003, doi: 10.1021/ci034160g.