SINERGI *Vol. 23, No.1, pp.29-37, April 2025*
DOI :http://dx.doi.org/10.31963/sinergi.v23i1.5507

Prediksi Kesehatan Mesin Diesel Menggunakan Metode SVM dan BPNN Berbasis Reduksi Ketidakseimbangan Data

Fadli Nurdin [1*], Mohammad Khoirul Effendi [1]

[1] Department Teknik Mesin, Institut Teknologi Sepuluh Nopember, Surabaya, Jawa Timur 60111, Indonesia
[*]Email: fadli.nurdin01@gmail.com

Abstract*: In predictive maintenance systems, class imbalance often affects the performance of machine learning models, particularly in accurately detecting faulty engine conditions. This study proposes a relabeling strategy using K-Means clustering to address class imbalance in diesel engine health prediction. The dataset, containing six operational parameters (engine RPM, oil pressure, fuel pressure, coolant pressure, oil temperature, and coolant temperature), was processed through outlier removal, correlation analysis, and class adjustment by shifting a portion of the majority class to the minority class. Two predictive models were developed: Support Vector Machine (SVM) with RBF and Sigmoid kernels, and Back Propagation Neural Network (BPNN) with [3 3 1] architecture using satlins activation. Evaluation results show that the relabeling process significantly improved model accuracy from 72.30% to 91.93%. BPNN achieved the highest performance with 94.21% accuracy, surpassing both SVM models and prior benchmark studies. The proposed data adjustment approach enhances prediction reliability by balancing class distribution, particularly in fault detection tasks.*

Keywords*: diesel engine health; SVM; BPNN; class imbalance; relabeling; clustering*

Abstrak: Dalam sistem pemeliharaan prediktif, ketidakseimbangan kelas seringkali memengaruhi kinerja model pembelajaran mesin, terutama dalam mendeteksi kondisi mesin yang tidak sehat. Penelitian ini mengusulkan strategi relabeling menggunakan algoritma K-Means clustering untuk mengatasi ketidakseimbangan kelas pada prediksi kesehatan mesin diesel. Dataset yang digunakan terdiri dari enam parameter operasional (RPM mesin, tekanan oli, tekanan bahan bakar, tekanan pendingin, suhu oli, dan suhu pendingin), yang diproses melalui tahap pembersihan data, analisis korelasi, serta penyesuaian distribusi label dengan memindahkan sebagian data dari kelas mayoritas ke minoritas. Dua model prediksi dikembangkan: Support Vector Machine (SVM) dengan kernel RBF dan Sigmoid, serta Back Propagation Neural Network (BPNN) dengan arsitektur [3 3 1] dan fungsi aktivasi satlins. Hasil evaluasi menunjukkan bahwa proses relabeling meningkatkan akurasi model secara signifikan dari 72,30% menjadi 91,93%. Model BPNN memberikan kinerja terbaik dengan akurasi 94,21%, melampaui model SVM maupun studi acuan sebelumnya. Pendekatan ini menunjukkan efektivitas dalam meningkatkan keandalan prediksi melalui distribusi data yang lebih seimbang, khususnya untuk mendeteksi kerusakan mesin.

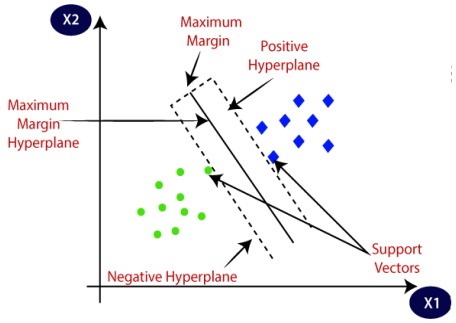
Kata kunci: kesehatan mesin diesel; SVM; BPNN; ketidakseimbangan data; relabeling; clustering

I. PENDAHULUAN

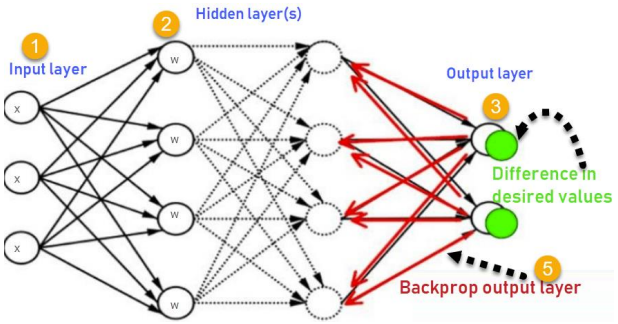
Dalam dunia industri modern, pendekatan *Condition-Based Maintenance* (CBM) semakin berkembang sebagai strategi utama untuk meningkatkan keandalan sistem dan menurunkan biaya pemeliharaan. CBM bergantung pada pemantauan parameter kondisi operasional mesin secara real-time guna memprediksi potensi kegagalan sebelum terjadi kerusakan fatal. Salah satu penerapan CBM yang penting adalah pada sistem mesin diesel penggerak utama kapal maupun kendaraan operasional, yang memiliki peran vital dan bekerja secara kontinu di bawah tekanan operasional tinggi.

Penelitian dalam bidang ini telah mengadopsi berbagai pendekatan pembelajaran mesin (*machine learning*) untuk memodelkan kesehatan mesin berdasarkan parameter seperti suhu, tekanan oli, dan putaran mesin. Metode seperti *Support Vector Machine* (SVM) dan *Back Propagation Neural Network* (BPNN) telah terbukti efektif dalam melakukan klasifikasi kondisi mesin baik atau rusak. SVM dikenal karena ketangguhannya dalam menangani data berdimensi tinggi dan margin pemisah yang optimal, sementara BPNN memiliki kemampuan dalam menangkap hubungan *non-linear* antar variabel yang

kompleks. Metode SVM dan BPNN dalam melakukan system prediksi ditunjukkan seperti pada gambar 1 dan 2.



Gambar 1. Cara kerja SVM



Gambar 2. Cara kerja BPNN

Namun, tantangan utama yang kerap muncul dalam praktik nyata adalah ketidakseimbangan kelas (*class imbalance*) dalam dataset historis. Pada banyak kasus, data mesin yang rusak (kondisi tidak sehat) jauh lebih sedikit dibandingkan data mesin yang masih sehat. Ketidakseimbangan ini menyebabkan model prediksi cenderung *bias* terhadap kelas mayoritas, sehingga mengurangi kemampuan deteksi terhadap kelas minoritas yang sebenarnya menjadi fokus utama dalam sistem pemeliharaan prediktif .

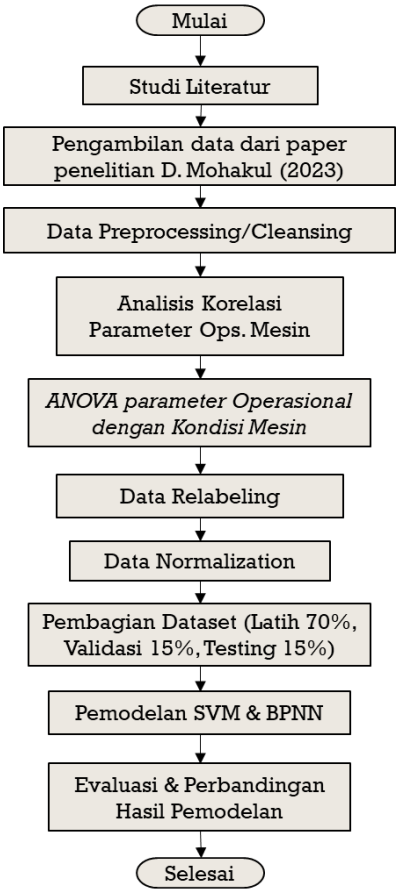
Salah satu pendekatan alternatif untuk mengatasi masalah ini adalah dengan menerapkan metode *relabeling* data. Relabeling dilakukan untuk merevisi label kelas berdasarkan struktur data internal menggunakan metode *unsupervised learning*. Dalam penelitian ini, digunakan algoritma K-Means Clustering untuk mempartisi ulang data ke dalam dua klaster berdasarkan kemiripan fitur, kemudian label baru diberikan pada klaster yang dominan mengandung data bermasalah. Validasi dilakukan dengan menggunakan *silhouette score* untuk menilai kualitas klaster yang terbentuk.

Beberapa penelitian terdahulu telah menunjukkan keberhasilan strategi relabeling. Zhang et al. [1] menerapkan pendekatan *dynamic relabeling* pada data fault diagnosis dan mendapatkan peningkatan akurasi model sebesar 10-15%. Demikian pula, Chen dan Wang [2] menyatakan bahwa integrasi antara *unsupervised relabeling* dan *supervised learning* mampu meningkatkan sensitivitas model terhadap data minoritas. Akan tetapi, implementasi relabeling pada kasus kesehatan mesin diesel masih jarang ditemukan, terutama dengan kombinasi evaluasi *silhouette score* dan *confusion matrix* secara sistematik.

Dengan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis dampak dari proses relabeling dataset terhadap akurasi prediksi kesehatan mesin diesel menggunakan metode SVM dan BPNN. Penelitian ini diharapkan dapat menjadi referensi praktis dalam mengembangkan sistem klasifikasi kesehatan mesin berbasis data yang lebih andal di lingkungan industri otomotif maupun kelautan.

II. METODE PENELITIAN

Penelitian ini dilaksanakan sebagai studi kuantitatif berbasis data eksperimental dari mesin diesel operasional. Tahapan penelitian disusun secara sistematis dan divisualisasikan dalam Gambar 1.

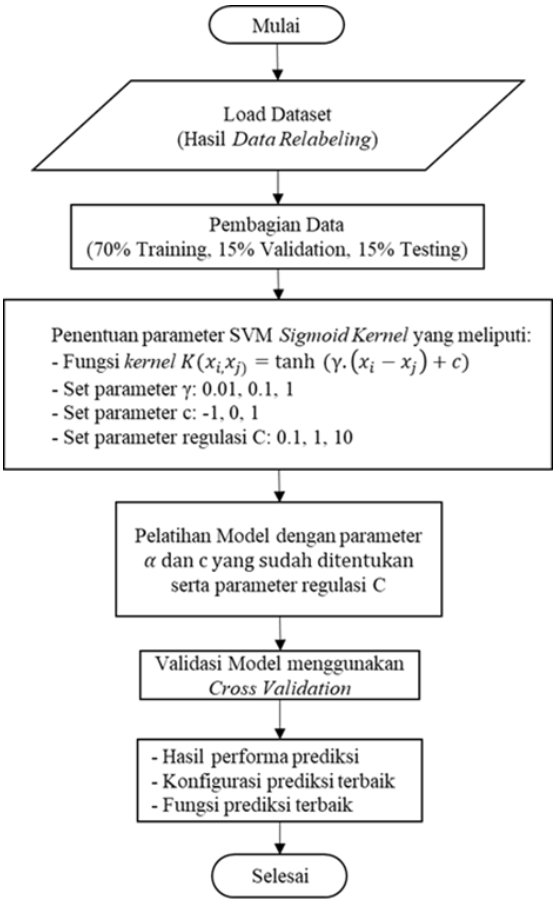


Gambar 3. Flowchart Metode Penelitian

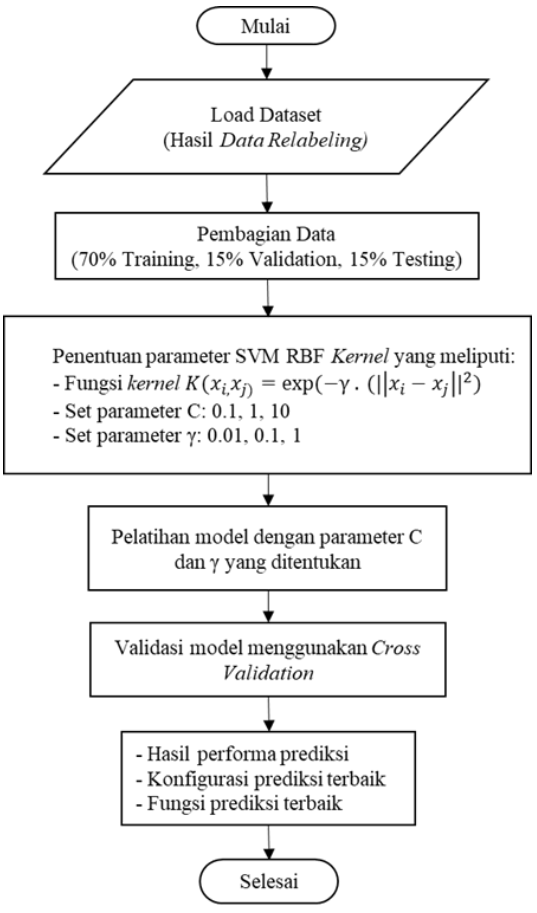
Penjelasan tahapan penelitian yang dilakukan sebagai berikut:

1. Akuisisi dan Preprocessing Data
   Dataset awal berisi 174 baris data asli dengan 6 parameter input *(Engine RPM, Lubrication Oil Pressure, Fuel Pressure, Coolant Pressure, Lubrication Oil Temperature, Coolant Temperature)* dan satu label output engine condition (kondisi mesin) "1" mesin sehat dan "0" mesin tidak sehat. Dataset merupakan data operasional mesin diesel *MTU Series 1400* penggerak utama kapal. Data diakuisisi melalui sensor-sensor mesin saat mesin dijalankan. Status mesin sehat (1) jika seluruh parameter berada pada standar operasional dan kondisi mesin saat beroperasi tidak ada gangguan seperti mengalami getaran berlebih, mesin mati, atau pincang. Sedangkan mesin tidak sehat (0) jika kondisi sebaliknya.
2. Korelasi dan Seleksi Fitur
   Analisis korelasi dilakukan menggunakan Spearman's Rank untuk menilai hubungan antar parameter, dilanjutkan dengan ANOVA deviance untuk mengidentifikasi parameter dominan terhadap kondisi mesin.
3. Deteksi *Outlier*
   Outlier dideteksi menggunakan metode *Interquartile Range* (IQR), dengan batas bawah dan atas ditentukan dari Q1 - 1.5*IQR dan Q3 + 1.5*IQR.

4. Data Relabeling (Reduksi Ketidakseimbangan Data)
   Dataset dilabel ulang menggunakan metode K-Means dengan jumlah klaster k=2. Setiap titik data dievaluasi ulang keanggotaannya terhadap klaster, dan relabeling dilakukan dengan memindahkan sebagian data dari klaster mayoritas ke minoritas dalam rentang eksperimen 5% hingga 50%. Evaluasi menggunakan nilai *silhouette score* untuk menjaga kualitas pemisahan klaster, serta akurasi prediksi sementara menggunakan BPNN.
5. Normalisasi dan Pembagian Data
   Seluruh parameter dinormalisasi dengan metode *Min-Max Scaling*. Dataset dibagi menjadi 70% untuk pelatihan, 15% validasi, dan 15% pengujian.
6. Pemodelan
   Dua pendekatan digunakan:
   - SVM dengan kernel RBF dan Sigmoid, dengan parameter masing-masing kernel diatur seperti parameter C, gamma, dan konstanta c. Proses pemodelan SVM dijelaskan pada diagram alir pemodelan pada gambar 2 dan 3 berikut.

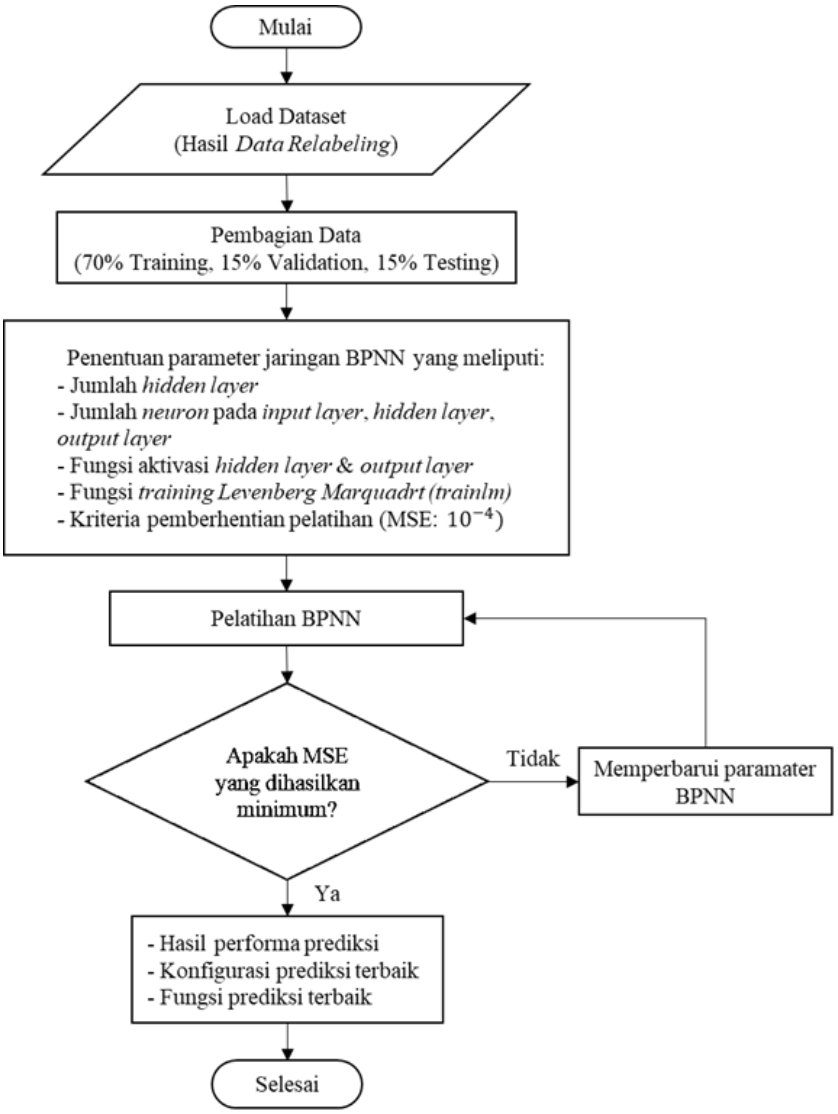


Gambar 4. Diagram alir pemodelan SVM Kernel RBF



Gambar 5. Diagram alir pemodelan SVM Kernel Sigmoid

- BPNN menggunakan menggunakan konfigurasi 1-*3 hidden layer*, dengan 1-3 neuron per layer, variasi fungsi aktivasi pada *hidden layer* dan *output layer*, serta algoritma pembelajaran *Levenberg-Marquardt (trainlm)*. Proses pemodelan BPNN dijelaskan pada diagram alir pemodelan pada gambar 4 berikut.



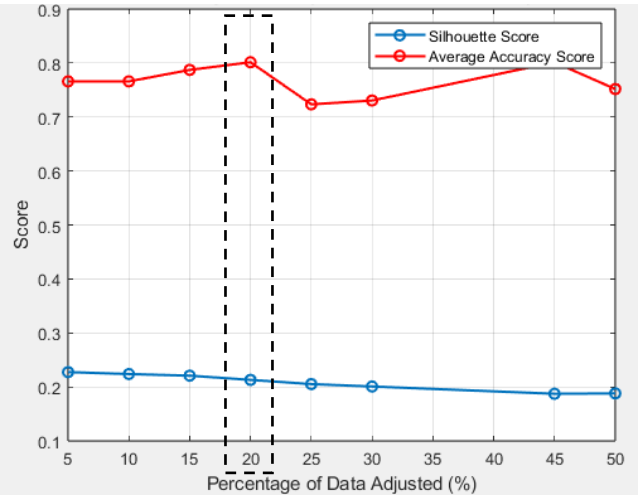
Gambar 3. Diagram alir pemodelan BPNN

7. Evaluasi Model
Model dievaluasi menggunakan metrik: akurasi, *precision, recall, F1-score.*

III. HASIL DAN PEMBAHASAN

3.1 Hasil Relabeling dengan Reduksi Ketidakseimbangan Data

Untuk mengurangi ketidakseimbangan kelas yang terjadi akibat dominasi *cluster* mayoritas, dilakukan proses penyesuaian data dengan memindahkan sebagian kecil data dari *cluster* mayoritas ke *cluster* minoritas. Proses ini bertujuan untuk memperoleh distribusi kelas yang lebih seimbang, serta meningkatkan kinerja klasifikasi model. Rentang variasi proporsi perpindahan data yang digunakan adalah sebesar 5% hingga 50% dari total data di kelas mayoritas, dengan kenaikan bertahap 5%. Rentang ini ditentukan berdasarkan desain eksperimen peneliti, guna mengevaluasi sejauh mana perubahan komposisi label berdampak terhadap kualitas klaster dan performa model prediksi. Setiap variasi diuji menggunakan model sementara BPNN, dan hasilnya dievaluasi berdasarkan nilai *silhouette score* dan akurasi klasifikasi. Berdasarkan hasil eksperimen, diperoleh bahwa perpindahan sebesar 20% dari data kelas mayoritas memberikan hasil terbaik, dengan nilai silhouette score yang masih stabil dan akurasi prediksi tertinggi, seperti ditunjukkan pada Gambar 4.



Gambar 4. *Silhoutte Score* dan Akurasi Prediksi Data *Adjustement*

Pendekatan ini sejalan dengan metode eksplorasi proporsi relabeling yang juga pernah digunakan oleh (Zhang et al., 2019) dalam studi klasifikasi berbasis imbalance adjustment, meskipun tanpa batas baku, pemilihan rentang proporsi sering kali ditentukan melalui percobaan berulang untuk menemukan titik optimum. Kemudian, studi oleh Putra et al. (2023) membahas penggunaan teknik resampling seperti SMOTE dan SMOTE-ENN dalam mengatasi ketidakseimbangan data. Evaluasi hasil klasterisasi dilakukan menggunakan silhouette score yaitu nilai seberapa baik proses clustering mempertahankan struktur dari data asli untuk menilai kualitas pengelompokan setelah penerapan metode resampling. Selain itu, penelitian oleh Rachman dan Susanto (2022) mengembangkan model prediksi berbasis clustering dan ensemble learning dalam pemrosesan data sensor untuk prediksi kegagalan sistem teknik. Pendekatan ini menunjukkan bahwa kombinasi antara teknik clustering dan metode ensemble learning dapat meningkatkan akurasi prediksi dengan mempertimbangkan parameter silhouette score (Rachman & Susanto, 2022). Penyesuaian ini pada akhirnya menghasilkan distribusi cluster akhir yang lebih seimbang, sebagaimana ditunjukkan pada Tabel 1

Tabel 1. Komposisi Data Setelah Relabeling Clustering K-Means dengan Adjust Data Kelas Mayoritas

Cluster	Data Awal	Data Relabeling
Engine Condition “0”	61	81
Engine Condition “1”	113	77

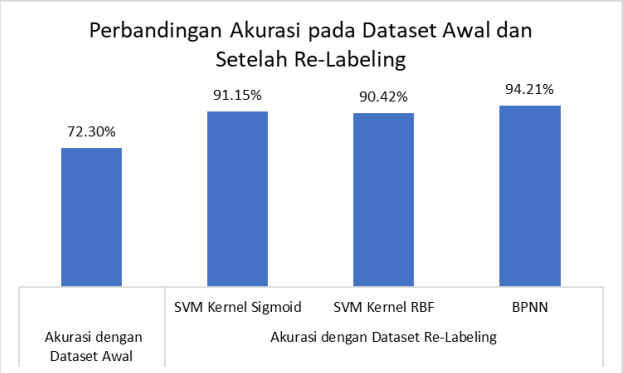
Kriteria pemetaan didasarkan pada rata-rata pembobotan berdasarkan nilai korelasi nilai fitur pada masing-masing cluster.

Distribusi data setelah pelabelan ulang dievaluasi untuk memastikan bahwa jumlah data pada kedua kategori mendekati ideal atau seimbang. Keseimbangan komposisi data dihitung dengan formula jumlah data masing-masing label kondisi mesin dibagi dengan jumlah keseluruhan data. Hasil evaluasi menunjukkan bahwa rasio komposisi setelah relabeling adalah 0.95, cukup menggambarkan keseimbangan data yang lebih baik dari sebelumnya 0.53. Hal ini menunjukkan bahwa optimasi distribusi cluster berhasil meningkatkan keseimbangan data.

3.2 Hasil Prediksi

Hasil penelitian menunjukkan bahwa proses reduksi ketidakseimbangan data pada proses *relabeling* dataset berdampak signifikan terhadap peningkatan performa model prediksi seperti

ditunjukkan pada gambar 4 perbandingan akurasi menggunakan dataset awal dan *relabeling*.



Gambar 4. Perbandingan Nilai Akurasi Model Setelah *Relabeling*

Sebelum *relabeling*, seluruh model menghasilkan akurasi <75%. Setelah *relabeling* 10%, akurasi meningkat drastis, khususnya pada model BPNN. Confusion matrix menunjukkan penurunan false negative secara signifikan. Untuk menilai performa model yang digunakan dalam penelitian ini, dilakukan perbandingan antara hasil prediksi dengan hasil penelitian referensi yang dilakukan oleh Mohakul et al. (2023) yang terdapat pada tabel 2 berikut.

Tabel 2. Perbandingan Hasil Prediksi SVM dan BPNN dengan penelitian referensi

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)	Sumber
SVM Sigmoid Kernel	91.15	99.26	89.93	94.17	Hasil penelitian
SVM RVF Kernel	90.42	88.29	93.64	90.85	Hasil penelitian
BPNN	94.21	93.61	95.37	94.22	Hasil penelitian
Logistic Regression	89.00	88.00	87.00	88.00	D. Mohakul, 2023
SVM Linear	88.75	88.00	87.00	88.00	D. Mohakul, 2023
SVM Polinomial Kernel	88.00	87.00	84.00	86.00	D. Mohakul, 2023
KNN	89.00	88.00	87.00	87.00	D. Mohakul, 2023
Naive Bayes	83.00	82.00	79.00	80.00	D. Mohakul, 2023
Decicion tree	74.00	72.00	67.00	68.00	D. Mohakul, 2023

Berdasarkan tabel perbandingan hasil prediksi, dapat dilihat bahwa metode BPNN yang digunakan dalam penelitian ini menunjukkan kinerja terbaik dengan akurasi sebesar 94.21%, presisi 93.61%, recall 95.37%, dan F1-score 94.22%. Sementara itu, metode SVM dengan kernel sigmoid yang digunakan dalam penelitian ini juga memberikan hasil yang lebih baik dibandingkan dengan model SVM yang digunakan dalam penelitian Mohakul (2023), khususnya dalam aspek akurasi 91.15% vs. 88.75% dengan SVM linear. Hal ini menunjukkan bahwa model yang digunakan dalam penelitian ini lebih optimal dalam memprediksi kondisi kesehatan mesin dibandingkan dengan model SVM yang digunakan dalam studi referensi.

3.3 Analisis Ekonomi Implementasi Sistem Prediksi

Selain meningkatkan akurasi dalam mendeteksi kondisi mesin, penerapan sistem prediksi kesehatan mesin diesel berbasis SVM dan BPNN ini juga memiliki implikasi ekonomi yang signifikan. Dengan deteksi dini terhadap ketidaknormalan mesin, sistem ini memungkinkan pengambilan keputusan pemeliharaan secara lebih tepat waktu, sehingga dapat mengurangi risiko kerusakan fatal dan biaya

overhoul yang mahal. Sebagai ilustrasi, berdasarkan data estimasi biaya pemeliharaan di lapangan, kerusakan besar pada mesin diesel tipe MTU Series 1400 dapat menimbulkan biaya hingga Rp. 150.000.000 per kejadian. Estimasi ini didasarkan pada pengalaman praktis dan perbandingan dengan katalog harga suku cadang mesin serta biaya tenaga kerja overhaul di lingkungan industri maritim dan otomotif.Dengan asumsi sistem ini mampu mengurangi satu kejadian kerusakan besar dalam satu siklus operasional (misalnya setiap 6 bulan), maka potensi penghematan langsung dapat mencapai nilai tersebut. Selain itu, dengan peningkatan keandalan deteksi kondisi mesin tidak sehat, perusahaan dapat merencanakan *downtime* dengan lebih efisien dan menghindari *downtime* tak terduga yang berdampak pada terganggunya operasional. Dengan demikian, sistem prediksi ini tidak hanya memberikan manfaat teknis, tetapi juga menjadi investasi strategis untuk efisiensi biaya jangka panjang.

3.4 Evaluasi Ketahanan Model

Evaluasi terhadap ketahanan model prediksi (Model Evaluasi Algoritma/MEA) perlu dilakukan untuk memastikan bahwa model tidak hanya berkinerja tinggi dalam data uji saat ini, tetapi juga stabil dalam jangka panjang saat digunakan secara berkelanjutan. Dalam studi ini, stabilitas model dievaluasi dengan mengamati fluktuasi performa prediksi selama proses pelatihan dan validasi, serta nilai *mean squared error* (MSE) yang rendah dan konsisten pada BPNN di angka rata-rata 0.21.

Namun, untuk implementasi nyata, model perlu dievaluasi dalam skenario siklus jangka panjang (*long term cyclic evaluation*), seperti variabilitas data operasional musiman, keausan komponen mesin seiring waktu, dan pergantian operator. Oleh karena itu, studi lanjutan direkomendasikan untuk melakukan pembelajaran ulang model secara berkala *(periodic retraining)* dengan data terbaru, serta mengintegrasikan sistem monitoring online untuk mengidentifikasi degradasi performa prediksi. Pendekatan ini akan menjaga validitas model dalam jangka panjang dan memperkuat keberlanjutan sistem prediktif di lingkungan operasional nyata.

IV. KESIMPULAN

Kesimpulan dari hasil penelitian dijelaskan pada bagian ini

1. Modifikasi SVM dengan kernel Sigmoid dan parameter C = 10, gamma = 0.1, dan c = 1 menghasilkan akurasi sebesar 91.15%. Meskipun cukup tinggi, model ini masih lemah dalam mendeteksi mesin tidak sehat, terbukti dari nilai *False Negative* yang tinggi.
2. Model BPNN dengan konfigurasi [3 3 1] dan fungsi aktivasi *satlins* menunjukkan kinerja terbaik, mencapai akurasi 94.21%. BPNN terbukti lebih efektif dalam menangkap pola operasional dan mendeteksi kondisi mesin tidak sehat.
3. BPNN lebih akurat dan konsisten dibanding SVM serta model lain dalam penelitian sebelumnya, menjadikannya metode yang lebih dapat diandalkan untuk prediksi kesehatan mesin diesel.
4. Proses *relabeling* dataset dengan K-Means dan *imbalance class adjustment* berhasil mengatasi ketidakseimbangan data. Hasilnya, akurasi rata-rata seluruh model meningkat signifikan dari 72.30% menjadi 91.93%, serta mengurangi bias klasifikasi terhadap kelas mayoritas.

DAFTAR PUSTAKA

[1] J. Zhang, L. Wang, dan Y. Chen, “Dynamic Relabeling for Imbalanced Fault Diagnosis Dataset Using Clustering and Classification,” IEEE Access, vol. 8, pp. 94321–94332, 2020, doi: 10.1109/ACCESS.2020.2994999.

[2] L. Chen dan B. Wang, “Improved Health Status Classification for Rotating Machinery via Cluster-Based Relabeling,” Mechanical Systems and Signal Processing, vol. 154, p. 107527, 2021, doi: 10.1016/j.ymssp.2020.107527.

[3] D. Park, H. Lee, dan Y. Kim, “A Hybrid Data Relabeling Method for Machinery Fault Classification Based on Unsupervised Clustering and Supervised Learning,” Applied Sciences, vol.

11, no. 12, p. 5440, 2021, doi: 10.3390/app11125440.
[4] B. Liu, Y. Zhao, dan D. Tang, "Random Forest-Based Predictive Maintenance Model for Equipment Condition Monitoring," Journal of Intelligent Manufacturing, vol. 28, no. 6, pp. 1307–1321, 2017, doi: 10.1007/s10845-015-1063-3.
[5] V. Vapnik, "The Nature of Statistical Learning Theory," Springer, New York, 1995.
[6] D. Mohakul, "Condition Based Predictive Maintenance on Ship's Major Equipment Using AI," IOP Conference Series: Materials Science and Engineering, vol. 1240, 2023.
[7] Y. LeCun, Y. Bengio, dan G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436–444, 2015, doi: 10.1038/nature14539.
[8] A. Adugna, A. Getachew, dan D. Tessema, "Support Vector Machines in Fault Diagnosis: A Review," International Journal of Engineering Research and Technology, vol. 11, no. 3, pp. 477–484, 2022.
[9] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, dan R. X. Gao, "Deep Learning and Its Applications to Machine Health Monitoring," Mechanical Systems and Signal Processing, vol. 115, pp. 213–237, 2019, doi: 10.1016/j.ymssp.2018.05.050.
[10] S. Doganis, A. Alexandridis, P. Patrinos, dan H. Sarimveis, "Time Series Sales Forecasting for Short Shelf-Life Food Products Based on Artificial Neural Networks and Evolutionary Computation," Journal of Food Engineering, vol. 75, pp. 196–204, 2006.