

KLASIFIKASI JENIS KEGAGALAN MESIN MENGUNAKAN HYPERPARAMETER TUNNING SVM DAN LOGISTIC REGRESSION

Ibnu Alfarobi¹, Sofian Wirahadi²,
Kudiantoro Widiyanto³,

Program Studi Teknologi Komputer¹

Program Studi Sistem Informasi²

Program Teknologi Informasi³

Fakultas Teknik dan Informatika¹²³

Universitas Bina Sarana Informatika¹²³

Email : ibnu.iba@bsi.ac.id, sofian.sod@bsi.ac.id

kudiantoro.kdw@bsi.ac.id

Received: May 10, 2023. **Revised:** June 15, 2023. **Accepted:** July 3, 2023. **Issue Period:** Vol.7 No.3 (2023), Pp.854-861

Abstrak: kegagalan peralatan atau mesin dari suatu proses manufaktur sering mengakibatkan kerugian finansial yang besar bagi bisnis. Berdasarkan tingkat jenis kegagalan mesin, model dengan penurunan kinerja adalah sama untuk berbagai jenis tetapi akan berbeda untuk faktor manusia, mesin dan lingkungan. Dengan mengevaluasi efek dan penanganan yang berbeda, menjadi sangat penting untuk dapat memprediksi faktor – faktor yang menentukan jenis kegagalan mesin. Dengan memanfaatkan model machine learning, pengklasifikasi diharapkan mampu memprediksi suatu jenis kegagalan dari mesin. Pada penelitian ini, memaksimalkan untuk mendapatkan hasil model dengan memanfaatkan hyperparameter yang di tuning pada algoritma Support Vector Machine dan Logistic Regression. Metode penanganan data seperti teknik Smote dan data preprocessing lainnya digunakan. Model menghasilkan perbedaan tingkat akurasi yang cukup besar yaitu 23%. Sehingga dapat disimpulkan model dengan menggunakan algoritma SVM bisa lebih baik dalam memprediksi jenis kegagalan mesin pada tingkat keakuratan sebesar 90% dibandingkan dengan model menggunakan algoritma Logistic Regression yang hanya menghasilkan keakuratan sebesar 67%.

Kata Kunci : SMOTE, Kegagalan Mesin, Support Vector Machine, Logistic Regression

Abstract: The failure of equipment or machinery of a manufacturing process often results in large financial losses for the business. Based on the degree of engine failure type, models with performance degradation are the same for different types but will be different for human, machine and environmental factors. By evaluating the different effects and handling, it becomes very important to be able to predict the factors that determine the type of engine failure. By utilizing machine learning models, classifiers are expected to be able to predict a type of failure from the machine. In this study, maximizing to get model results by utilizing hyperparameters tuned in the Support Vector Machine and Logistic Regression algorithms. Data



DOI: 10.52362/jisamar.v7i3.771

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

handling methods such as Smote techniques and other data preprocessing are used. The model produces a considerable difference in accuracy rate of 23%. So it can be concluded that models using the SVM algorithm can be better at predicting the type of machine failure at an accuracy level of 90% compared to models using the Logistic Regression algorithm which only produces accuracy of 67%.

Keywords: SMOTE, Machine Failure, Support Vector Machine, Logistic Regression

I. PENDAHULUAN

Industri manufaktur jadi lini prioritas dalam implementasi industri 4. 0, kalau pelaksanaan teknologi di bidang industri manufaktur dalam melaksanakan analisis buat memprediksi pemeliharaan perlengkapan serta mesin penciptaan secara pas serta responsif. Strategi serta teknologi predictive maintenance menolong industri memperoleh keunggulan kompetitif. Masa gunakan perlengkapan pembedahan bisa dioptimalkan, pengurangan bayaran perawatan, pengurangan downtimes, kenaikan pengetahuan, kenaikan keselamatan kerja, kenaikan keandalan pabrik industri, pemanfaatan sumber energi perawatan yang efektif, pelaksanaan strategi tanpa cacat, serta perencanaan proses yang fleksibel serta dimaksimalkan yang sanggup berkontribusi pada keunggulan kompetitif jangka panjang [1].

Sistem mekanis memburuk bersamaan waktu, oleh sebab itu pemeliharaan dibutuhkan buat membenarkan keandalan predictive maintenance bertujuan buat kurangi bayaran serta revisi yang memakan waktu, serta pula menjauhi kegiatan yang tidak butuh menganjurkan strategi pemeliharaan yang diinformasikan oleh mesin pemantauan keadaan [2].

Kegagalan perlengkapan serta waktu henti yang pendek dari proses manufaktur kerap menimbulkan kerugian ekonomi yang besar untuk industri. Bersumber pada tingkatan keparahan degradasi kinerja mesin serta penggalian pengetahuan dari sebagian input, model penyusutan kinerja dari bermacam tipe perlengkapan yang sama, namun aspek manusia, mesin, serta area yang berbeda. Dengan penilaian daya guna perawatan pada kebijakan perawatan yang berbeda, bayaran relevan, sumber energi, serta strategi perawatan yang maksimal bisa didetetapkan [3].

Riset yang berfokus pada predictive maintenance dengan pendekatan analisis klasifikasi bersumber pada perbandingan karakteristik mesin yang hadapi ‘kegagalan mesin’ dalam baris informasi spesial ini ataupun tidak hadapi kegagalan mesin. Dengan menggunakan bermacam model machine learning, pengklasifikasi diharapkan sanggup memprediksi keadaan mesin apakah kandas mesin ataupun tidak kandas mesin sehingga bisa jadi input dalam mengestimasi bayaran pemeliharaan dan kenaikan masa gunakan mesin penciptaan [4].

Bersamaan dengan pertumbuhan ilmu pengetahuan serta teknologi data, kedatangan cabang ilmu baru di bidang pc informasi mining sudah menarik banyak atensi dalam dunia sistem data [5].

Klasifikasi ialah proses informasi mining yang bertujuan buat membagi informasi ke dalam kelas- kelas buat mempermudah pengambilan keputusan [6].

Algoritma ataupun tata cara klasifikasi sangat bermacam- macam, salah satunya merupakan Support Vector Machine(SVM). SVM merupakan sesuatu mesin linier dengan sebagian watak yang sangat baik, ilham utama dari SVM merupakan buat membangun sesuatu hyperline selaku pengambil keputusan sedemikian rupa sehingga margin pembelahan antara kelas positif serta negatif dioptimalkan [6].

Algoritma Support Vector Machine mempunyai kelebihan antara lain merupakan dalam memastikan jarak memakai Support Vector Machine sehingga proses komputasi jadi kilat. Ada pula kekurangan Support Vector Machine(SVM) merupakan susah dipakai problem berskala besar [7].



Logistic Regression ialah klasifikasi linier yang sudah teruji menciptakan klasifikasi yang powerful dengan statistik probabilitas serta bisa menanggulangi permasalahan klasifikasi multi kelas. Kelemahan dari Logistic Regression merupakan rentan terhadap underfitting pada dataset yang kelasnya tidak balance, sehingga mempunyai akurasi yang rendah [8].

Buat menanggulangi kasus dataset tidak balance merupakan dengan menyeimbangkan distribusi kelas tidak seragam di antara kelas- kelas memakai tata cara SMOTE serta biar jumlahnya balance dari kelas kebanyakan ataupun kelas minoritas [9]

Tata cara klasifikasi yang digunakan buat mencari nilai akurasi yang memakai machine learning, setelah itu dataset tersebut diuji serta di penilaian yang hendak diseleksi tata cara yang menciptakan nilai akurasi sangat besar [10] [11].

II. METODE DAN MATERI

1. Data Exploration

Data *exploration* dilakukan dengan tujuan untuk mendapatkan pemahaman tentang data yang digunakan. Yaitu dengan cara melakukan pengecekan dimensi data, isi data, *type* data, *missing values*, *duplicated* data, unik data yang menjadi *feature* target model klasifikasi dengan nama *feature* 'Failure Type', distribusi jumlah data berdasarkan label kelas dari *feature* 'Failure Type'. Kemudian melakukan pengecekan terkait data *outlier* pada *features* Air temperature [K], Process temperature [K], Rotational speed [rpm], Torque [Nm], dan Tool wear [min]. *Feature* 'Failure Type' merupakan label kelas dari suatu jenis kegagalan mesin yang terjadi, dan *feature* 'Type' merupakan sebagai varian kualitas produk dan nomor seri khusus varian.

2. Data Preprocessing

Data *preprocessing* yaitu melakukan penanganan dengan menggunakan metode - metode *machine learning* untuk mengatasi permasalahan – permasalahan yang ada pada data sehingga mendapatkan hasil model klasifikasi dengan tingkat akurasi yang lebih baik. Yaitu melakukan *features selection* dengan menghapus *features* UDI dan Product ID karena *feature* UDI merupakan hanya Unique Id dan *feature* Product ID merupakan ID product pada sebuah mesin yang tidak dapat memberikan informasi penting terhadap suatu jenis kegagalan mesin. Sehingga jumlah *feature* yang digunakan untuk membuat model menjadi 7 *features* diantaranya *features* Type, Air temperature [K], Process temperature [K], Rotational speed [rpm], Torque [Nm], Tool wear [min], dan Failure Type. Kemudian melakukan *label encoder* pada *features* yang bertipe data *string* yaitu 'Type' dan 'Failure Type' menggunakan fungsi *Label Encoder* dari *scikit learn*. Untuk menangani data *outlier* dilakukan penghapusan data *outlier* menggunakan rumus statistik $IQR=Q3-Q1$. Pembagian data dilakukan dengan persentase 80:20 yaitu 80% menjadi data train dan 20% menjadi data test. Data train digunakan untuk melatih model klasifikasi machine learning, dan data test digunakan untuk evaluasi model klasifikasi. Data input pada data train dan data test dilakukan *feature scaling* menggunakan fungsi *MinMaxScaler* dari *scikit learn*, dengan tujuan membuat data menjadi skala standar dengan menskalakan setiap fitur ke rentang nilai antara 0 dan 1.

3. Smote

Smote (*Synthetic Minority Oversampling Technique*) digunakan untuk mengatasi ketidak seimbangan kelas yang terjadi pada data dengan membuat data sintesis pada kelas minoritas yang mengikuti berdasarkan jumlah data pada kelas mayoritas. Smote yang dilakukan hanya pada data train sebelum dilakukannya *training* model klasifikasi.

4. Hyperparameter Tunning



Algoritma *Support Vector Machine* yaitu menggunakan *hyperparameter* kernel, gamma, dan nilai C. Ketiga *hyperparameter* tersebut di *tunning* dengan beberapa nilai menggunakan GridsearchCV. Nilai dari *hyperparameter* tersebut disajikan pada Tabel 1. dibawah ini:

Tabel 1. Hyperparameter SVM

Hyperparameter	Nilai
Kernel	RBF dan Linear
Gamma	1, 0.1, dan 0.01
C	1, 10, dan 100

Kemudian algoritma *Logistic Regression* yaitu menggunakan *hyperparameter* penalty, dan *hyperparameter* C. Kedua *hyperparameter* tersebut di *tunning* dengan beberapa nilai menggunakan GridsearchCV. Nilai dari *hyperparameter* tersebut disajikan pada Tabel 2. dibawah ini:

Tabel 2. Hyperparameter Logistic Regression

Hyperparameter	Nilai
Penalty	L1 dan L2
C	1, 10, dan 100

5. *Support Vector Machine*

Algoritma *Support Vector Machine* digunakan untuk membuat model klasifikasi machine learning yang dapat memprediksi sebuah jenis kegagalan mesin. Model menggunakan data train yang sudah dilakukan teknik smote. Model di training dengan 10-fold cross validation pada refit True, dan verbose 2. Model dibuat dengan melakukan penambahan metode OneVsOneClassifier. Metode ini bekerja berdasarkan bahwa pada dasarnya algoritma support vector machine yang hanya khusus digunakan untuk klasifikasi dua kelas, kemudian pada saat digunakan pada klasifikasi multiclass maka metode OneVsOneClassifier melakukan proses training dengan metode heuristic yaitu membangun sejumlah model SVM biner yang membandingkan satu kelas dengan kelas lainnya. Terakhir model dievaluasi menggunakan data test yang sebesar 20%.

6. *Logistic Regression*

Algoritma *Logistic Regression* digunakan untuk membuat model klasifikasi machine learning yang dapat memprediksi sebuah jenis kegagalan mesin . Model menggunakan data train yang sudah dilakukan teknik smote. Model di training dengan 10-fold cross validation pada refit True, dan verbose 2. Model dibuat dengan melakukan penambahan metode OneVsOneClassifier. Terakhir model dievaluasi menggunakan data test yang sebesar 20%.

III. PEMBAHASA DAN HASIL

1. Hasil Data *Exploration*

Berdasarkan hasil dari melakukan data *exploration* terdapat sebanyak 10 *features* dan 10.000 *instances*, serta tidak adanya indikasi data yang mengandung nilai *uplicated* dan *missing values*. Kemudian dari hasil pengecekan isi data dan tipe data telah terdapat adanya kesesuaian antara isi data dengan jenis tipe data dari setiap *features*. Yaitu 3 *features* bertipe data *string* yaitu Product ID, Type, dan Failure Type. 3 *features* bertipe data *float* yaitu Air temperature [K], Process temperature [K], dan Torque [Nm]. 4 *features* bertipe data *interger*



DOI: 10.52362/jisamar.v7i3.771

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

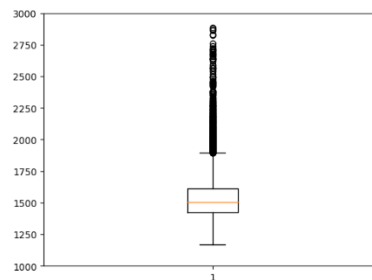
yaitu UDI, Rotational speed [rpm], Tool wear [min], dan Target.

Jumlah data berdasarkan label kelas target model klasifikasi ditunjukkan pada Tabel 3. Terindikasi adanya jumlah data yang tidak seimbang, artinya dimana pada satu kelas terdapat jumlah data yang mayoritas yaitu sebanyak 9.652 data dengan label No Failure atau data yang tidak diidentifikasi jenis kegagalannya dan pada label lain terdapat jumlah data yang minoritas yaitu sebanyak 18 data dengan label Random Failures.

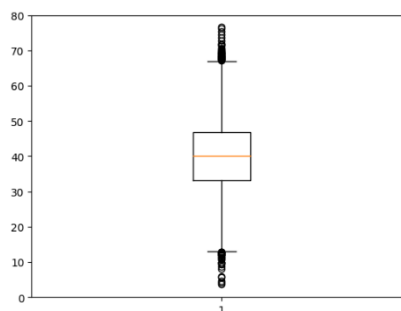
Tabel 3. Feature Failure Type

Label	Jumlah
Heat Dissipation Failure	112
No Failure	9.652
Overstrain Failure	78
Power Failure	95
Random Failures	18
Tool Wear Failure	45

Data *outlier* pada pengecekan kelima *features* tersebut, telah terindikasi adanya data *outlier* pada *features* Rotational Speed [rpm] Gambar 1 dan Torque [Nm] pada gambar 2.



Gambar 1. Indikasi Data Outlier feature Rotational Speed [rpm]



Gambar 2. Indikasi Data Outlier feature Torque [Nm]

2. Hasil Data Preprocessing

Hasil *label encoder* dari feature ‘Type’ dan ‘Failure Type’ masing – masing disajikan pada Tabel 4. dan Tabel 5.

Tabel 4. Label Encoder Feature Type

Sebelum Label Encoder	Sesudah Label Encoder
H	0



DOI: 10.52362/jisamar.v7i3.771

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

L	1
M	2

Tabel 5. Label Encoder Feature Failure Type

Sebelum Label Encoder	Sesudah Label Encoder
Heat Dissipation Failure	0
No Failure	1
Overstrain Failure	2
Power Failure	3
Random Failures	4
Tool Wear Failure	5

Setelah dilakukan penghapusan data outlier sebanyak 465 data, jumlah data menjadi berkurang sehingga jumlah data menjadi 9.535 data. Jumlah distribusi data berdasarkan label feature Failure Type pada Tabel 6. dibawah ini. Dan hasil pembagian data yaitu data train sebanyak 7.628 data, dan data test sebanyak 1.907 data.

Tabel 6. Feature Failure Type

Label	Jumlah
Heat Dissipation Failure	111
No Failure	9.223
Overstrain Failure	78
Power Failure	64
Random Failures	18
Tool Wear Failure	41

3. Hasil Smote

Tabel 7. merupakan Hasil dari pada proses melakukan teknik Smote pada data train sebelum training model klasifikasi multiclass data dengan label yang sedikit yaitu label 0 (Heat Dissipation Failure) sebanyak 93 data, 2 (Overstrain Failure) 56 data, 3 (Power Failure) 49 data, 4 (Random Failures) sebanyak 16, dan 5 (Tool Wear Failure) 33 data, distribusi menjadi sebanyak 7.381 data pada setiap masing – masing label mengikuti jumlah distribusi data terbanyak dengan label 1 (No Failure), sehingga jumlah data input dan output model data training menjadi sebanyak 44.286 data.

Tabel 7. Hasil Smote

Label	Sebelum Smote	Setelah Smote
Heat Dissipation Failure (0)	93	7.381
No Failure (1)	7.381	7.381
Overstrain Failure (2)	56	7.381
Power Failure (3)	49	7.381
Random Failures (4)	16	7.381
Tool Wear Failure (5)	33	7.381

4. Hasil Evaluasi Model SVM

Hasil evaluasi model klasifikasi machine larning dengan menggunakan algoritma support vector machine untuk memprediksi jenis kegagalan mesin dengan label data multiclass, model terbaik yang dihasilkan yaitu pada kernel RBF, nilai C=100, dan nilai gamma=1 dengan nilai best score untuk training sebesar 99%. Dari hasil nilai



DOI: 10.52362/jisamar.v7i3.771

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

hyperparameter terbaik tersebut kemudian digunakan untuk mengevaluasi model klasifikasi machine learning menggunakan data test yang sudah dilakukan sampai pada tahap feature scalling. Hasil evaluasi model mendapatkan nilai akurasi sebesar 90%, f1-score 58%, precision 56%, dan recall 64%. Dimana pada f1-score, precision dan recall menggunakan pembobotan everage macro.

5. Hasil Evaluasi Model *Logistic Regression*

Hasil evaluasi model klasifikasi machine larning dengan menggunakan algoritma logistic regression untuk memprediksi jenis kegagalan mesin dengan label data multiclass, model menghasilkan nilai hyperparameter terbaik pada algoritma logistic regression untuk klasifikasi multiclass yaitu pada nilai $C=100$, dan nilai $\text{Penalty}=L2$ dengan nilai best score untuk training sebesar 88%. Dari hasil nilai hyperparameter terbaik tersebut yang akan digunakan untuk mengevaluasi model machine learning menggunakan data test yang sudah dilakukan sampai pada tahap feature scalling. Hasil evaluasi model mendapatkan nilai akurasi sebesar 67%, f1-score 46%, precision 42%, dan recall 71%. Dimana pada f1-score, precision dan recall menggunakan pembobotan everage macro.

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan, bahwa diantara algoritma support vector machine dan logistic regression yang digunakan untuk membuat model klasifikasi machine learning dalam memprediksi suatu jenis kegagalan mesin, model terbaik yang dihasilkan yaitu ketika pada saat menggunakan algoritma support vector machine karena menghasilkan tingkat akurasi paling tinggi yaitu sebesar 90% jika dibandingkan dengan logistic regression yang hanya menghasilkan akurasi sebesar 67%, dimana selisi ini cukup besar yaitu 23%. Kemudian hasil model terbaik ini dapat dipertanggung jawabkan karena menggunakan average macro yang tidak dapat mengabaikan kelas minoritas pada data yang diprediksi yaitu data test dengan score precision sebesar 56%, recall 64%, dan f1-score 58%. Dengan hasil ini model SVM mampu mengenerate data multiclass dengan baik.

REFERENASI

- [1] A. Bousdekis, D. Apostolou, and G. Mentzas, "Predictive Maintenance in the 4th Industrial Revolution: Benefits, Business Opportunities, and Managerial Implications," *IEEE Eng. Manag. Rev.*, vol. 48, no. 1, pp. 57–62, 2020.
- [2] S. Kearthland and T. L. Van Zyl, "Automating predictive maintenance using oil analysis and machine learning," *2020 Int. SAUPEC/RobMech/PRASA Conf. SAUPEC/RobMech/PRASA 2020*, 2020.
- [3] J. Yan, Z. Zhang, L. Xie, and Z. Zhu, "A unified framework for decision tree on continuous attributes," *IEEE Access*, vol. 7, pp. 11924–11933, 2019.
- [4] K. Y. Nazara, "Perancangan Smart Predictive Maintenance untuk Mesin Produksi," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 691–702, 2022.
- [5] I. Irmawati, K. Widiyanto, F. Aziz, A. Rifai, and A. Rahmawati, "Implementasi artificial neural network dalam mendeteksi penyakit hati (liver)," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 6, no. 1, pp. 193–198, 2022.
- [6] Y. Alkhalifi, A. Fazriansyah, M. S. Azis, and K. Widiyanto, "Deep learning untuk pendeteksian penyakit kanker payudara dengan optimasi Adam," vol. 7, no. 1, pp. 124–136, 2023.
- [7] A. Desiani, M. Akbar, Irmeilyana, and A. Amran, "Implementasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular," *J. Tek. Elektro dan Komputasi*, vol. 4, no. 2, pp. 207–214, 2022.
- [8] H. Rianto and R. S. Wahono, "Resampling Logistic Regression untuk Penanganan Ketidakseimbangan Class pada Prediksi Cacat Software," *J. Softw. Eng.*, vol. 1, no. 1, pp. 46–53, 2015.
- [9] W. Feng, W. Huang, and W. Bao, "Imbalanced Hyperspectral Image Classification with an Adaptive Ensemble Method Based on SMOTE and Rotation Forest with Differentiated Sampling Rates," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1879–1883, 2019.
- [10] B. Jonathan, P. H. Putra, and Y. Ruldeviyani, "Observation Imbalanced Data Text to Predict Users



- Selling Products on Female Daily with SMOTE, Tomek, and SMOTE-Tomek,” *Proc. - 2020 IEEE Int. Conf. Ind. 4.0, Artif. Intell. Commun. Technol. IAICT 2020*, pp. 81–85, 2020.
- [11] F. Y. Pamuji, S. Dwi, and A. Putri, “Komparasi Metode SMOTE dan ADASYN Untuk Penanganan Data Tidak Seimbang MultiClass,” pp. 331–338, 2021.



DOI: 10.52362/jisamar.v7i3.771

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).