

DETEKSI DINI KERUSAKAN BEARING MENGGUNAKAN *MACHINE LEARNING* PENDEKATAN *SUPPORT VECTOR REGRESSION* (SVR)

*Endhy Priambodo¹, Toni Prahasto², Achmad Widodo³

¹Mahasiswa Jurusan Teknik Mesin, Fakultas Teknik, Universitas Diponegoro

²Dosen Jurusan Teknik Mesin, Fakultas Teknik, Universitas Diponegoro

³Dosen Jurusan Teknik Mesin, Fakultas Teknik, Universitas Diponegoro

Jl. Prof. Sudharto, SH., Tembalang-Semarang 50275, Telp. +62247460059

*E-mail: endhy28@gmail.com

Abstrak

Bearing adalah sebuah komponen mekanis yang digunakan untuk mengurangi gesekan antara dua komponen yang bergerak relatif satu sama lain. Bantalan biasanya terdiri dari sebuah lingkaran luar dan lingkaran dalam yang dipisahkan oleh benda bulat yang disebut bola atau rol. Bantalan dapat digunakan pada berbagai jenis mesin dan peralatan, seperti mesin industri, kendaraan, dan alat berat. Kondisi yang buruk pada bantalan dapat menyebabkan kerusakan pada mesin dan mengurangi umur pemakaian mesin. Oleh karena itu, pemantauan dan perawatan yang tepat pada bantalan sangat penting untuk memastikan kinerja mesin yang optimal dan menghindari kerusakan yang tidak diinginkan. Dalam rangka mencegah terjadinya kerusakan yang lebih serius, maka dibutuhkan suatu proses yang dapat memprediksi potensi terjadinya kegagalan dan menentukan estimasi sisa waktu operasi atau *remaining useful life* (RUL) dari suatu komponen sebelum terjadi kegagalan. Penelitian ini bertujuan untuk menganalisis kinerja algoritma *machine learning* dalam memprediksi kondisi kerusakan bantalan NASA (*National Aeronautics and Space Administration*) menggunakan dataset NASA Bearing Data Set. Dataset ini kemudian diolah sehingga mempermudah dalam proses *machine learning*, serta dilakukan ekstraksi fitur. Setelah memilih fitur terbaik, langkah selanjutnya adalah melakukan beberapa tahapan untuk mempersiapkannya sebagai input untuk algoritma *machine learning Support Vector Regression* (SVR). Hasil penelitian menunjukkan terdapat 12 model yang dihasilkan dari proses tersebut dengan akurasi prediksi yang berbeda-beda namun mampu memberikan estimasi waktu kapan bantalan akan mengalami kegagalan, yang dihitung mulai dari saat terdeteksi adanya kerusakan. Dengan adanya model ini, diharapkan dapat membantu dalam melakukan prediksi kerusakan bantalan sehingga perawatan dan perbaikan dapat dilakukan lebih awal dan mencegah terjadinya kerusakan yang lebih parah.

Kata kunci: bearing nasa; *machine learning*; rul; statistic; svr

Abstract

Bearing, a mechanical component used to reduce friction between two moving parts, usually consists of an outer ring, an inner ring, and a ball or roller separating them. Bearings can be used in various types of machines and equipment, such as industrial machinery, vehicles, and heavy equipment. Poor bearing conditions can lead to machine damage and reduce its service life. Therefore, proper monitoring and maintenance of bearings are essential to ensure optimal machine performance and avoid unwanted damage. In order to prevent more serious damage, a process is needed to predict the potential failure and estimate the remaining useful life (RUL) of a component before failure occurs. This research aims to analyze the performance of machine learning algorithms in predicting the condition of NASA (National Aeronautics and Space Administration) bearings using the NASA Bearing Data Set. The dataset is processed to simplify the machine learning process and feature extraction is performed. After selecting the best features, several steps are taken to prepare them as input for the Support Vector Regression (SVR) machine learning algorithm. The research results in 20 models with varying prediction accuracies but capable of providing an estimate of when the bearing will fail, calculated from the detection of damage. With this model, it is expected to help predict bearing damage so that maintenance and repair can be carried out earlier, preventing more serious damage.

Keywords: bearing nasa; *machine learning*; rul; statistic; svr

1. Pendahuluan

Bearing merupakan salah satu komponen penting dalam mesin yang berfungsi untuk menopang dan memposisikan poros secara radial pada saat berputar. Kinerja dan efisiensi bearing sangat mempengaruhi keberhasilan pengoperasian sistem dan mekanisme pada sebuah alat [1]. Kegagalan pada *bearing* dalam suatu mesin merupakan salah satu jenis kegagalan yang paling sering dijumpai. Salah satunya pada kegagalan pada *asynchronous motor* dimana 45-55% disebabkan oleh adanya kegagalan pada komponen bearing [2]. Kegagalan pada *bearing* diantaranya terjadi pada bagian *outer-race*, *inner-race*, *roller*, dan *cage*. Kegagalan pada bearing dapat menyebabkan malfungsi atau kerusakan lainnya pada sistem mesin yang berakibat pada terjadinya *downtime* yang lama dan *maintenance cost* yang tinggi [3]. Hal inilah yang mendorong dilakukannya proses prognostik untuk memprediksi potensi terjadinya kegagalan dan menentukan estimasi dari sisa masa efektif komponen untuk beroperasi sebelum terjadi kegagalan (*remaining useful life*). Prognostik sendiri merupakan bagian dari disiplin ilmu baru yaitu *Prognostics Health Management* (PHM) yang juga menjadi strategi baru dalam bidang *maintenance* mesin maupun komponen mesin [4].

Analisis prognostik, yang juga dikenal sebagai deteksi kerusakan dini, memiliki tiga aspek penting yang perlu diperhatikan. Pertama, penentuan atau penyusunan indikator kesehatan komponen, yang bertujuan untuk mengevaluasi kondisi sistem mesin dan komponen kritisnya saat ini [5]. Hal ini digunakan sebagai dasar untuk memprediksi sisa masa manfaat (RUL). Kedua, prediksi RUL digunakan untuk memperkirakan berapa lama komponen dapat digunakan sebelum mengalami kegagalan. Ketiga, manajemen kesehatan mesin atau komponen digunakan untuk mengoptimalkan kinerja sistem mesin dengan memantau kondisi komponen secara terus-menerus [6].

Dalam analisis prognostik khususnya pada bearing, data merupakan elemen penting yang digunakan untuk mengidentifikasi kondisi komponen dan melakukan analisis [7]. Data tersebut diperoleh melalui sensor yang dipasang pada alat dan mencatat berbagai variabel dengan interval waktu tertentu. Data kemudian diolah dan disimpan dalam bentuk log atau langsung dikirim ke sistem pengawasan (monitoring) [8]. Terdapat berbagai jenis sensor yang digunakan seperti sensor getaran, suhu, kimia, emisi akustik, dan tekanan suara. Sebelum data dari sensor tersebut digunakan untuk analisis, perlu dilakukan proses pengolahan seperti filtering dan penghilangan noise. Tidak semua data yang diperoleh dari sensor pada bearing digunakan dalam analisis prognostik. Hanya data yang paling signifikan dalam mengindikasikan terjadinya *fault* pada bearing yang dipilih. Proses ini dikenal sebagai *feature selection*. Sebelum melakukan analisis diagnosis dan prognosis, data dari sensor tersebut harus melalui tahap *preprocessing data* terlebih dahulu. Tahap ini melibatkan pengolahan data untuk memastikan kualitas data yang digunakan dalam analisis.

Untuk melakukan analisis prognostik yang efektif, langkah penting yang harus dilakukan adalah memilih algoritme yang sesuai dengan properti sistem dan karakteristik data, serta menerapkan algoritme tersebut dengan benar untuk mendapatkan hasil yang optimal [5]. Oleh karena itu, diperlukan kriteria pemilihan algoritme yang terukur dan juga sistem penilaian yang akurat, seperti *benchmarking*. Dalam penelitian ini, beberapa pendekatan dan algoritme akan digunakan pada data yang sama untuk menentukan hasil analisis yang paling optimal, dan hasil dari tiap pendekatan akan dibandingkan dengan menggunakan kriteria akurasi. Data yang telah melalui tahap *preprocessing* akan dibagi menjadi dua bagian yaitu *train data* dan *test data*, dengan rasio tertentu. *Train data* akan digunakan untuk menyusun model sesuai dengan algoritme yang dipilih, sedangkan *test data* akan digunakan untuk mengukur keberhasilan dari model tersebut terhadap data baru. Selain itu, *train data* dan *test data* harus terpisah dan tidak ada data yang sama antara keduanya.

Analisis prognostik menggunakan *Support Vector Regression* (SVR) telah digunakan pada dataset bearing NASA. Metode ini memungkinkan untuk melakukan prediksi sisa masa manfaat (RUL) pada bearing dengan menggunakan data yang dihasilkan dari sensor getaran. Data tersebut telah melalui proses *preprocessing* sebelum dianalisis menggunakan metode SVR [10]. Pada dataset ini, SVR digunakan untuk menghasilkan model prediksi RUL yang akurat dan dapat diandalkan. Hasil dari analisis menggunakan metode SVR menunjukkan bahwa model tersebut mampu menghasilkan prediksi RUL dengan tingkat akurasi yang tinggi. Selain itu, metode SVR juga memungkinkan untuk melakukan analisis fitur sehingga dapat dipilih fitur-fitur yang paling penting dalam melakukan prediksi RUL pada bearing. Analisis prognostik dengan menggunakan metode SVR pada dataset bearing NASA dapat memberikan informasi yang sangat berharga dalam mengurangi biaya perawatan dan meningkatkan efisiensi dari sistem pemeliharaan mesin.

Adapun tujuan dari tugas akhir ini adalah sebagai berikut:

1. Analisis dilakukan untuk mengevaluasi bagaimana pendekatan algoritme *machine learning Support Vector Regression* (SVR) dapat digunakan dalam melakukan analisis prognostik pada dataset bearing NASA dengan tujuan memprediksi estimasi sisa masa manfaat (*Remaining Useful Life*).

2. Mengetahui bagaimana perbedaan metode pengolahan getaran dalam ekstraksi fitur dari data bearing dapat mempengaruhi performa model SVR dalam memprediksi *Remaining Useful Life* (RUL) pada dataset bearing NASA.
3. Mengevaluasi kinerja model yang dihasilkan menggunakan pendekatan algoritme *Support Vector Regression* (SVR) serta menentukan parameter model yang paling baik dalam memprediksi *Remaining Useful Life* (RUL) pada dataset bearing NASA.

2. Bahan dan Metode Penelitian

2.1. Pengumpulan Data Getaran NASA Bearing Dataset

Dalam melakukan analisis prognosis yaitu melakukan prediksi untuk nilai *Remaining Useful Life* (RUL) dari bearing NASA, digunakan kumpulan data yang diperoleh secara daring yang dipublikasikan oleh NASA dan pusat Pusat *NSF I/UCR Intelligent Maintenance Systems* (IMS). Data ini diambil dari 4 buah bearing yang terinstalasi dengan poros. Kecepatan putar pada bearing dilakukan dengan konstan pada angka 2000 RPM. Data bearing NASA terbagi menjadi 3 set berdasarkan waktu pengambilan data serta kuantitas dalam pengambilan data.

Tabel 1. Spesifikasi NASA bearing dataset

Set 1	Waktu pengambilan data	22 Oktober 2003 - 25 November 2003
	Jumlah data	2156
Set 2	Waktu pengambilan data	12 Februari 2004 – 19 Februari 2004
	Jumlah data	984
Set 3	Waktu pengambilan data	4 Maret 2004 – 18 April 2004
	Jumlah data	4448

2.2. Preprocessing Raw Data dan EDA (*Exploratory Data Analysis*)

Preprocessing raw data pada dataset bantalan NASA melibatkan serangkaian langkah untuk mempersiapkan data mentah yang diperoleh dari pengukuran bantalan mesin oleh NASA untuk analisis lebih lanjut. Tujuan utama dari *preprocessing* adalah menghilangkan gangguan, meningkatkan kualitas data, mengurangi kebisingan, dan mempersiapkan data dalam format yang sesuai untuk analisis lebih lanjut. Data mentah harus diimpor ke lingkungan pemrograman seperti Python. Data ini biasanya berupa file teks atau file CSV yang berisi pengukuran sensor bearing NASA. Data ini harus dipahami terlebih dahulu dengan memeriksa struktur, format, dan jenis data yang terkandung di dalamnya.

Langkah berikutnya adalah membersihkan data dari entri yang tidak valid atau hilang. Ini melibatkan identifikasi dan penghapusan data duplikat, penggantian nilai-nilai yang hilang atau kosong, dan penanganan nilai ekstrem atau outlier yang tidak realistis. Selanjutnya data yang telah dibersihkan dilakukan *Exploratory Data Analysis*. Dimana *Exploratory Data Analysis* (EDA) merupakan proses untuk memahami, menganalisis, dan mengeksplorasi dataset bantalan NASA secara rinci. EDA bertujuan untuk menemukan pola, hubungan, anomali, dan wawasan menarik lainnya yang terkandung dalam data tersebut. Dalam konteks dataset bantalan NASA, EDA dapat memberikan pemahaman yang lebih baik tentang kondisi bantalan, performa mesin, dan faktor-faktor yang mempengaruhi kegagalan bantalan.

2.3. Ekstraksi Fitur dari Sinyal Getaran

Setelah dilakukan *pre-processing* data dari format (.mat), dilakukan ekstraksi fitur untuk merepresentasikan kondisi bantalan pada analisis ini. Metode yang dipilih adalah metode *time domain*, karena dapat mengidentifikasi ciri khas sinyal berbasis waktu seperti sinyal getaran dan/atau emisi akustik, yang dapat dibandingkan terhadap kondisi bantalan saat mengalami kecatatan. Dari beberapa metode *time domain* yang tersedia, dipilihlah analisis berdasarkan perhitungan statistik untuk menghasilkan tren dari spektrum yang telah dihasilkan, sehingga dapat dianalisis pola yang mampu mendeteksi adanya cacat bantalan yang baru muncul. Selain itu, data getaran dari bearing NASA juga dianalisis secara stationer dengan pendekatan *time domain* menggunakan metode perhitungan statistik. Hasil perhitungan statistik yang diperoleh langsung dari data getaran bearing NASA, diakumulasi dalam periode harian sesuai hari pengambilan data. Terdapat 12 jenis perhitungan statistik yang digunakan pada data getaran bearing NASA.

Tabel 2. Daftar Fitur yang Diekstrak

Fitur	Rumus Perhitungan
Rata-Rata Absolut	$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i $
Standar Deviasi (STD)	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$
Skewness	$SK = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - \bar{x})^3}{\sigma^3}$
Kurtosis	$K = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - \bar{x})^4}{\sigma^4}$
Entropi	$H(X) = - \sum_{i=1}^N P(x_i) \log P(x_i)$
Root Mean Square (RMS)	$x_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N x^2}$
Maksimum Absolut	$x_{max} = \max value $
Peak to Peak (P2P)	$x_p = \max value - \min value$
Faktor Puncak	$x_{crest} = \frac{\max value}{x_{rms}}$
Clearance Factor	$x_{clear} = \frac{x_{rms}}{(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i })^2}$
Shape Factor	$Shape\ factor = \frac{x_{rms}}{\bar{x}}$
Impuls	$Impulse = \frac{\max value}{\bar{x}}$

2.4. Membuat Set-Time Features

Pembuatan *set time feature* pada dataset bantalan NASA melibatkan langkah-langkah untuk mengambil informasi waktu dari data mentah yang berkaitan dengan pengukuran sensor bantalan. Tujuannya adalah untuk melihat pola dan tren yang terjadi seiring waktu pada data bantalan. Dengan dilakukannya *set time feature*, data lebih mudah terbaca karena dilakukan pengelompokan perhari dari banyaknya data yang diambil dalam satu hari. Dalam penulisan coding dilakukan dengan metode `.groupby()` adalah sebuah fungsi yang digunakan untuk mengelompokkan data berdasarkan nilai-nilai pada kolom atau kolom-kolom tertentu dalam suatu DataFrame. Fungsi ini berasal dari pustaka Pandas, yang sering digunakan dalam analisis data. Dengan begitu analisis pada data NASA bearing dapat dilakukan lebih mudah dan terfokuskan.

2.5. Pemilihan Dataset dari Data Getaran Bearing

Dalam pemilihan dataset dari NASA Bearing Dataset, salah satu faktor yang penting adalah jumlah data yang tersedia. Dataset yang memiliki jumlah data yang lebih banyak cenderung memberikan informasi yang lebih akurat dan representatif tentang kondisi bantalan mesin yang diamati. Dengan jumlah data yang lebih banyak, kita dapat memiliki sampel yang lebih besar untuk menganalisis pola, tren, dan hubungan antara variabel dalam dataset tersebut. Selain itu, dengan jumlah data yang lebih besar, kita dapat memperoleh hasil analisis yang lebih konsisten dan reliabel. Semakin banyak data yang digunakan dalam analisis, semakin tinggi pula kepercayaan statistik dan akurasi hasil yang diperoleh.

Dengan demikian, memilih Dataset 3 dari NASA Bearing Dataset yang memiliki jumlah data yang paling banyak merupakan langkah yang tepat untuk mengoptimalkan analisis dan pemahaman terhadap performa dan kondisi bantalan mesin. Data yang diambil disimpan file dengan format (.mat) dan dibagi sesuai dengan hari pengambilan data yang nantinya akan dilakukan ekstraksi fitur dan disimpan kedalam file dengan format (.csv).

2.6. Pemilihan Fitur dari Data Getaran Bantalan

Fitur yang dipilih dalam analisis prognosis memiliki peran penting dalam menentukan tingkat akurasi dari analisis yang dilakukan. Jumlah fitur yang digunakan tidak selalu berbanding lurus dengan kinerja analisis, terutama dalam memprediksi nilai RUL, dimana semakin banyak fitur yang digunakan tidak selalu meningkatkan akurasi analisis atau

prediksi. Oleh karena itu, perlu dilakukan pemilihan fitur yang paling berpengaruh dalam melakukan analisis prognosis, yaitu fitur yang mampu menggambarkan terjadinya degradasi pada bearing NASA. Fitur ini disebut sebagai indikator kesehatan dari bearing NASA. Dalam memilih fitur terbaik, akan digunakan perhitungan metrik yang terdiri dari dua komponen metrik, yaitu *monotonicity* (Mon) untuk mengukur tren yang terjadi pada nilai fitur, dan *trendability* (Rob) untuk mengukur ketahanan nilai fitur terhadap outliers. Untuk menghitung metrik tersebut, fitur harus didekomposisi menjadi tren rata-rata dan bagian acak menggunakan metode pemulusan. Dengan persamaan $X(t_k) = X_T(t_k) + X_R(t_k)$. perhitungan nilai kebaikan fitur. Dalam analisis ini akan digunakan nilai $\omega_1 = 0.5$ untuk *monotonicity*, $\omega_2 = 0.3$ untuk *trendability*

Tabel 3. Rumus Perhitungan Metrik Kebaikan Fitur

Fitur	Rumus Perhitungan
<i>Monotonicity</i> (Mon)	$Mon(X) = \left \frac{\text{no. of } \frac{d}{dx} > 0}{n-1} - \frac{\text{no. of } \frac{d}{dx} < 0}{n-1} \right $
<i>Trendability</i> (Tren)	$Tren(X) = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$

2.7. Proses Smoothing pada Fitur Terbaik yang Dipilih

Dalam penelitian ini, fitur terbaik yang telah diekstrak dari data getaran bantalan NASA dipilih dengan menggunakan metrik-*monotonicity*, *trendability*. Namun, nilai pada fitur-fitur tersebut memiliki noise tinggi yang mengakibatkan tren pada fitur tidak terlihat secara jelas. Oleh karena itu, dilakukan proses smoothing atau penghalusan nilai fitur dengan menggunakan empat algoritma smoothing yang berbeda yaitu *Gaussian Smoothing*, *Exponential Moving Average*. Tujuan dari proses smoothing adalah untuk mengurangi noise pada fitur sehingga tren yang dibentuk oleh data pada masing-masing fitur dapat terlihat dengan jelas. Hasil dari proses smoothing ditampilkan dalam bentuk grafik untuk masing-masing nilai fitur yang ditampilkan bersamaan dengan nilai asli fitur. Proses *smoothing* sangat penting dalam memprediksi nilai RUL pada bearing NASA karena proses ini akan menentukan performa dari fungsi yang terbentuk dalam memprediksi nilai RUL.

2.8. Pembagian Data untuk Train dan Test

Pada analisis ini, seluruh data yang digunakan untuk fitur terbaik sebelum proses *smoothing* dan keempat metode *smoothing* dibagi menjadi dua bagian, yaitu data *train* dan *test*. Data *train* sebesar 40% dari total data digunakan untuk melatih model, dengan rentang waktu dari hari ke-1 hingga hari ke-20. Sementara data *test* sebesar 60% digunakan untuk menguji performa model, dengan rentang waktu dari hari ke-21 hingga hari ke-46. Tujuan dari pembagian data ini adalah untuk memastikan bahwa model yang dibangun dapat melakukan prediksi secara akurat pada data baru. Data yang termasuk dalam bagian *train* digunakan untuk mencari nilai koefisien untuk melakukan standarisasi nilai fitur dan juga koefisien untuk melakukan *Principal Component Analysis*. Data *train* dalam hal ini adalah data operasional lampau dari bearing NASA yang bekerja secara normal. Sedangkan data *test* merupakan data baru dari operasional bearing NASA yang tidak diketahui kondisi performa dari bearing NASA tersebut. Untuk kasus ini, data *test* merepresentasikan performa dari kerja bearing NASA yang sedang diuji.

2.9. Standarisasi Nilai Fitur Terbaik yang Dipilih

Sebelum melanjutkan analisis lebih lanjut, fitur-fitur yang dipilih untuk prognosis dan pembuatan model perlu disesuaikan skala dan distandarisasi. Hal ini dilakukan karena banyak algoritma dalam machine learning yang dapat bekerja lebih baik ketika fitur berada pada skala yang relatif sama dan mendekati distribusi normal. Proses penyesuaian skala dilakukan dengan mengubah rentang nilai tanpa mengubah bentuk distribusi. Sementara proses standarisasi mengubah nilai menjadi standar deviasi distribusinya bernilai satu.

Untuk memperoleh hasil analisis yang akurat, perlu dilakukan proses standarisasi dan penyesuaian skala pada fitur-fitur yang dipilih sebelum dilakukan prognosis dan pembuatan model. Untuk melakukan proses tersebut, digunakan fungsi *StandardScaler* dari library *Scikit-learn*. Fungsi tersebut berguna untuk menghilangkan perbedaan rentang nilai pada fitur-fitur yang memiliki karakteristik input yang berbeda-beda. Dalam proses ini, mean akan dihapus dan data akan diubah ke varians unit. Hal ini akan mengatasi masalah ketika terdapat perbedaan rentang nilai pada fitur-fitur yang dipilih. Proses standarisasi dengan *StandardScaler* dilakukan pada data *train* untuk memperoleh koefisien, yang kemudian digunakan untuk melakukan penyesuaian skala dan standarisasi pada keenam fitur yang dipilih, baik untuk data *train* maupun *test*.

Tabel 4. Rumus Perhitungan Koefisien dari StandardScaler

Koefisien	Rumus Perhitungan
StandardScaler (Nilai Fitur Baru) Rata-rata (<i>Mean</i>)	$z = \frac{x_i - \mu}{\sigma} \times scale$ $Mean(\mu) = \frac{1}{N} \sum_{i=1}^N (x_i)$
Standar Deviasi (<i>Variances</i>)	$Variances(\sigma) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$
Skala (<i>Scale</i>)	$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$

2.10. Pembuatan Fitur dengan Principal Component Analysis (PCA)

PCA atau *Principal Component Analysis* adalah suatu teknik statistik non-parametrik yang tidak memerlukan label atau target yang digunakan untuk mengurangi dimensi dalam pembelajaran mesin. Apabila suatu dataset memiliki banyak fitur, maka disebut memiliki dimensi yang tinggi. Dimensi yang tinggi ini menjadi masalah dalam pembelajaran mesin karena model yang dihasilkan dapat mengalami *overfitting* sehingga kemampuan untuk menggeneralisasi data di luar contoh dalam set pelatihan menjadi menurun. Oleh karena itu, PCA dilakukan untuk mengurangi dimensi dalam suatu dataset, sehingga model yang dihasilkan menjadi lebih efisien, mempercepat pembelajaran, mengurangi biaya komputasi, dan meminimalkan fitur yang berlebihan. Selain itu, PCA juga dapat digunakan untuk memfilter *noise* pada kumpulan data seperti pada kompresi grafik. Setelah dilakukan PCA, data akan menghasilkan beberapa komponen dimana komponen utama mengekspresikan *varians* paling banyak dengan *noise* yang rendah, sementara komponen tambahan mengekspresikan *varians* yang lebih sedikit dengan *noise* yang lebih banyak.

Untuk menganalisis prognosis pada data getaran bearing NASA, digunakan enam fitur terbaik yang dipilih dari fitur-fitur yang berhasil diekstraksi dari data getaran bearing NASA. Kemudian, dilakukan standarisasi nilai dan proses PCA untuk menghasilkan dua komponen. Namun, hanya satu komponen PCA yang digunakan karena dianggap mampu merepresentasikan data asli dengan baik. Proses PCA dilakukan pada data *train* untuk memperoleh nilai koefisien PCA, yang kemudian digunakan untuk melakukan operasi dot pada ke enam fitur untuk menghasilkan satu fitur baru. Proses PCA dilakukan pada fitur yang tidak dilakukan proses *smoothing* dan enam fitur yang dilakukan proses *smoothing* dengan 2 metode. Hasil dari proses PCA pada seluruh variasi fitur ditampilkan dalam bentuk grafik untuk memudahkan dalam melihat nilai dari fitur hasil PCA.

2.11. Prediksi Nilai RUL dengan Metode Support Vector Regression (SVR)

Pada analisis data getaran bearing NASA, dilakukan ekstraksi fitur melalui beberapa perhitungan statistik, pemilihan fitur terbaik, standarisasi, penyesuaian skala, dan proses PCA. Kemudian, dibuat model untuk memprediksi nilai RUL bantalan menggunakan fitur-fitur yang dipilih sebagai input. Metode yang digunakan untuk pembuatan model adalah *Support Vector Regression* (SVR). Terdapat 3 variasi pembuatan model yang setiap variasinya dibagi menjadi 2 bagian. Tujuan dari analisis ini adalah untuk memprediksi nilai RUL dari bearing NASA.

Untuk mencari parameter terbaik dari algoritma SVR, digunakan *library* Optuna Dalam analisis ini parameter dari model SVR yang akan dioptimasi terangkum pada Tabel 4

Tabel 5. Parameter yang Dioptimasi pada Model SVR untuk Prediksi Nilai RUL Bearing NASA.

Parameter	Keterangan	Nilai
Kernel	Jenis kernel yang digunakan pada model untuk mempelajari data <i>train</i> dan melakukan prediksi	linear, poly, rbf, sigmoid
Degree	Derajat polynomial pada kernel poly, pada kernel lain nilainya diabaikan	1, 2, 3
Gamma Coef0	Nilai koefisien untuk kernel jenis rbf, poly, dan sigmoid Fungsi independen dalam kernel yang hanya signifikan pada kernel poly dan sigmoid	scale, auto 0.01 hingga 0.4, dengan langkah 0.01
Epsilon	Epsilon dalam model epsilon-SVR. Nilai Ini menentukan tabung epsilon di mana tidak ada penalti yang terkait fungsi kerugian pada proses <i>training</i> dengan poin yang diprediksi dalam jarak epsilon dari nilai aktual.	0.01 hingga 0.4, dengan langkah 0.01

Untuk menilai performa dari tiap model dalam melakukan prediksi nilai RUL, digunakan perhitungan skor dengan menggunakan *Root Mean Square Error* (RMSE) dan Koefisien Determinasi (R^2). Nilai RMSE sendiri mencerminkan

performa model dalam melakukan prediksi dimana semakin kecil nilai RMSE maka semakin baik performa dari model. Tabel 5 merangkum rumus perhitungan untuk RMSE dan R²

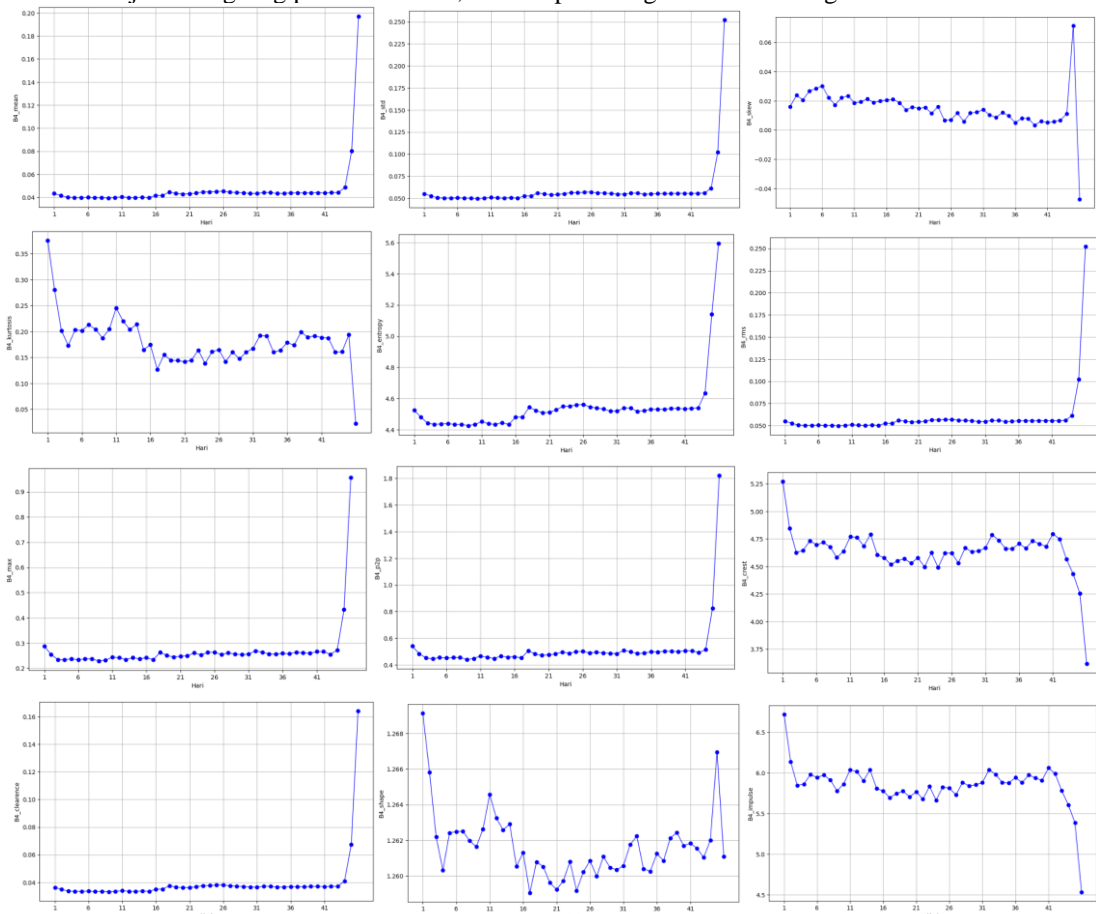
Tabel 6. Rumus Perhitungan *Root Mean Square Error* (RMSE) dan Koefisien Determinasi (R²)

Metrik	Rumus Perhitungan
<i>Root Mean Square Error</i> (RMSE)	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Prediksi_i - Aktual_i)^2}$
Koefisien Determinasi (R ²)	$R^2 = \left(\frac{N(\sum xy) - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}} \right)^2$

3. Hasil dan Pembahasan

3.1. Hasil Ekstraksi Data Getaran

Dari hasil seluruh perhitungan statistik yang dilakukan pada data vibrasi bearing, diperoleh hasil sebanyak 12 fitur dengan total 46 data untuk setiap fitur menurut jumlah hari pengambilan data. Hasil perhitungan statistik fitur dari data vibrasi bantalan disajikan langsung pada Gambar 1, dimana perhitungan statistik menghasilkan 12 fitur.



Gambar 1. Grafik hasil ekstraksi fitur dari data getaran bearing secara langsung.

3.2. Pemilihan Fitur Terbaik dari Hasil Ekstraksi Fitur pada Data Getaran Bantalan Secara Langsung dan dari Hasil *Spectral Kurtosis*

Proses ekstraksi fitur yang dilakukan pada data getaran bearing menghasilkan 12 fitur yang kemudian dilakukan pemilihan fitur dengan menggunakan kombinasi dari dua metrik yaitu *monotonicity*, dan *trendability* dengan nilai beban metrik yang telah ditentukan masing-masing. Tabel 6 menunjukkan hasil perhitungan fitur terbaik dan dari perhitungan tersebut dipilih enam yang terbaik untuk digunakan sebagai indikator kesehatan dari bearing.

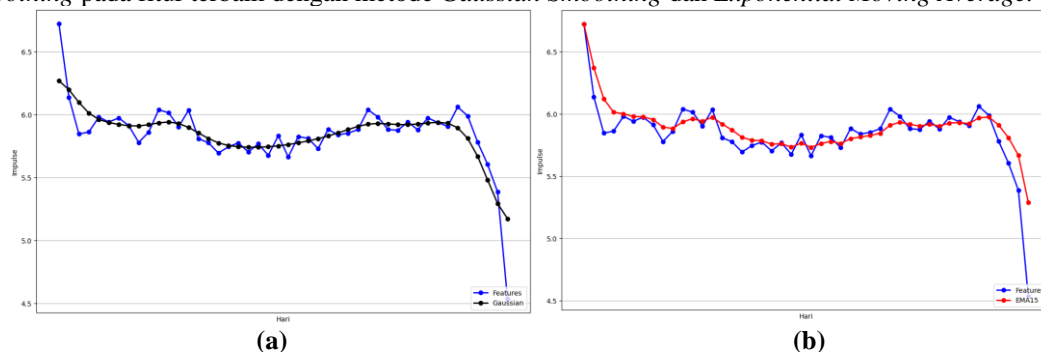
Tabel 7. Hasil Perhitungan Nilai Kebaikan Fitur Sebagai Indikator Kesehatan

Fitur	Monotonicity	Trendability	Goodness Metric
Impulse	0.048600	0.991748	0.520174
Faktor Puncak	0.036747	0.992142	0.514444
Entropi	0.025976	0.996074	0.511025
Peek to Peek	0.030711	0.969426	0.500068
Root Mean Square	0.000224	0.999586	0.499905
Maksimum Absolut	0.016178	0.968394	0.492286
Faktor Shape	0.004503	0.965527	0.485015
Standard Deviasi	0.004505	0.965462	0.484983
Rata-Rata Absolut	0.003501	0.966022	0.484761
Faktor Clearance	0.002911	0.966249	0.484580
Kurtosis	0.007844	0.951438	0.479641
Skewness	0.002640	0.874478	0.438559

Dari Tabel 7 yang menunjukkan hasil perhitungan kualitas fitur yang diekstraksi sebagai fitur untuk memprediksi nilai RUL bearing adalah enam fitur teratas, sehingga enam fitur terbaik yang dipilih untuk digunakan untuk analisis lebih lanjut adalah Impulse, Faktor Puncak, Entropi, Peek to Peek, Faktor Shape, dan Maksimum Absolut.

3.3. Data Smoothing pada Fitur Terbaik dari Data Getaran Bantalan

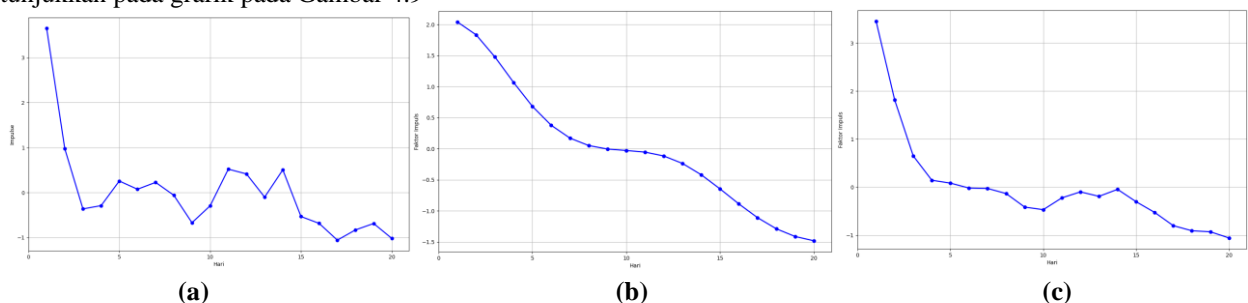
Proses *smoothing* dua metode untuk fitur terpilih terbaik menghasilkan nilai fitur dengan noise atau variasi lebih sedikit dari nilai aslinya. Hal ini terlihat dari grafik hasil proses *smoothing* yang menunjukkan kehalusan dari data getaran bearing yang lebih jelas pada grafik dibandingkan sebelum dilakukan *smoothing*. Pada Gambar 2. menunjukkan hasil proses *smoothing* pada fitur terbaik dengan metode *Gaussian Smoothing* dan *Exponential Moving Average*.



Gambar 2. Hasil Proses *Smoothing* pada Salah Satu Fitur Terbaik dengan Metode *Gaussian Smoothing* dan *Exponential Moving Average*,

3.4. Hasil Standarisasi Skala Terhadap Tiap Nilai Karakteristik

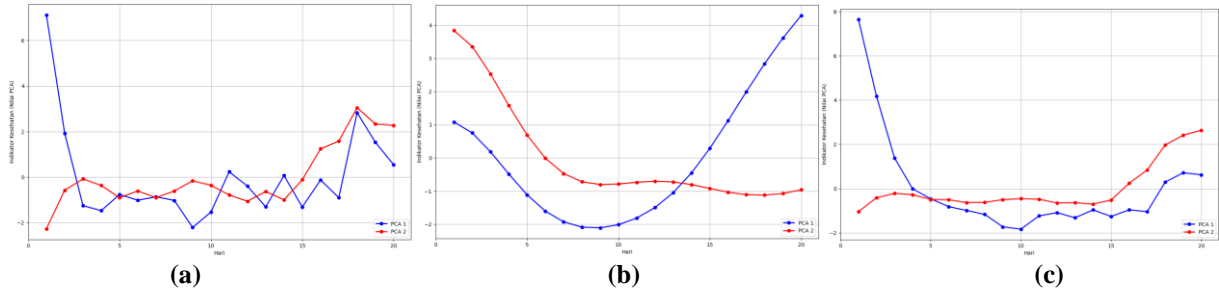
Penyesuaian standarisasi dan penskalaan dibuat untuk semua berbagai fitur yang digunakan sebagai input dalam membangun model untuk memprediksi nilai RUL inventaris. Untuk melakukan proses ini, fungsi *StandardScaler* dari pustaka *Pandas* digunakan. Pada data *train*, nilai dari enam fitur terbaik yang tidak dihaluskan, setelah dilakukan standarisasi dan penskalaan dengan perintah *StandardScaler*, akan mendapatkan nilai baru untuk setiap fitur yang ditunjukkan pada grafik pada Gambar 4.9



Gambar 3. Hasil Standarisasi Skala (a) Tanpa Proses *Smoothing* (b) *Gaussiang Smoothing*, (c) *Exponential Moving Average*.

3.5. Hasil Fitur dari Proses PCA dari Fitur Terbaik yang Dipilih

Setelah melakukan proses standarisasi pada enam fitur terbaik, dilakukan *Principal Component Analysis* (PCA) untuk menghasilkan fitur baru yang merepresentasikan fitur asli. Dalam analisis ini, dua komponen fitur dihasilkan, tetapi hanya satu fitur PCA yang dipilih sebagai input untuk model SVR. Pemilihan fitur PCA ini didasarkan pada visualisasi grafik pada data latih. Koefisien PCA yang diperoleh digunakan untuk menghasilkan komponen PCA dari data latih dan uji. Proses PCA dilakukan pada nilai fitur yang tidak dihaluskan dengan *smoothing*. Dari proses ini dihasilkan dua komponen baru yang mewakili data asli, yaitu nilai enam fitur terbaik yang dipilih dari sekian banyak fitur bagus yang diekstraksi langsung dari data getaran bearing. Sebelum proses PCA dijalankan, nilai fitur yang dipilih telah distandarisasi dan diskalakan agar setiap fitur yang digunakan memiliki bobot yang sama dalam membentuk nilai komponen baru pada proses PCA.



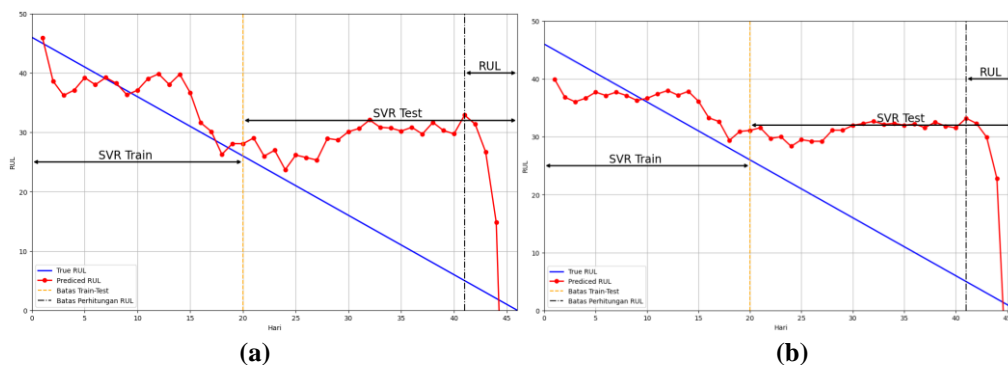
Gambar 4. Hasil pembuatan PCA (a) Tanpa Proses *Smoothing* (b) *Gaussian Smoothing*, (c) *Exponential Moving Average*.

3.6. Hasil Pembuatan Model SVR untuk Prediksi RUL

Model *Support Vector Regression* (SVR) dihasilkan dengan input enam fitur terbaik yang dipilih dari total 12 fitur yang berhasil diekstraksi dari data getaran bearing. Terdapat dua variasi utama yaitu model SVR tanpa PCA dan dengan PCA. Pada variasi tanpa menggunakan proses PCA, fitur pada delapan variasi dihaluskan untuk mereduksi noise pada nilai fitur dan pada kedua variasi tanpa menggunakan proses smoothing. Berbagai fitur distandarisasi dan diskalakan dengan *StandardScaler*. Pada tahap pertama dibuat model SVR dengan kernel linier dan tahap kedua dilakukan optimasi model dengan mencari parameter terbaik dari model. Kinerja model diukur dengan menghitung RMSE dan koefisien determinasi (R^2).

3.6.1. Model SVR Tanpa *Smoothing*

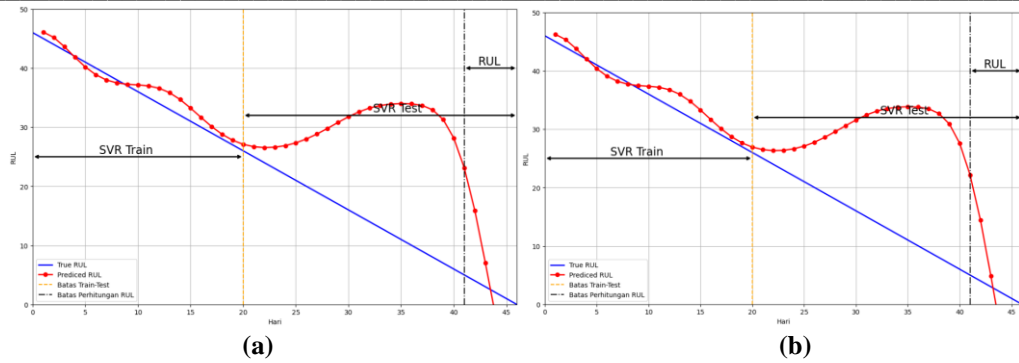
Prediksi RUL oleh model dengan input tanpa proses *smoothing* ditampilkan pada Gambar 5. Pada model dengan kernel linier, model berhasil mencapai nilai RUL= 0 pada hari ke 44 sehingga model memprediksi bearing memiliki RUL selama 3 hari. Sedangkan pada model optuna hasil prediksi menyentuh nilai RUL = 0 pada hari ke 44. Dari hasil tersebut, model yang dioptimalkan memprediksi bahwa bantalan akan memiliki Nilai dari 3 hari. Hasil keduanya memiliki kemiripan.



Gambar 5. Hasil prediksi nilai RUL tanpa proses *smoothing* menggunakan (a) kernel linier (b) optuna

3.6.2. Model SVR *Gaussian Smoothing*

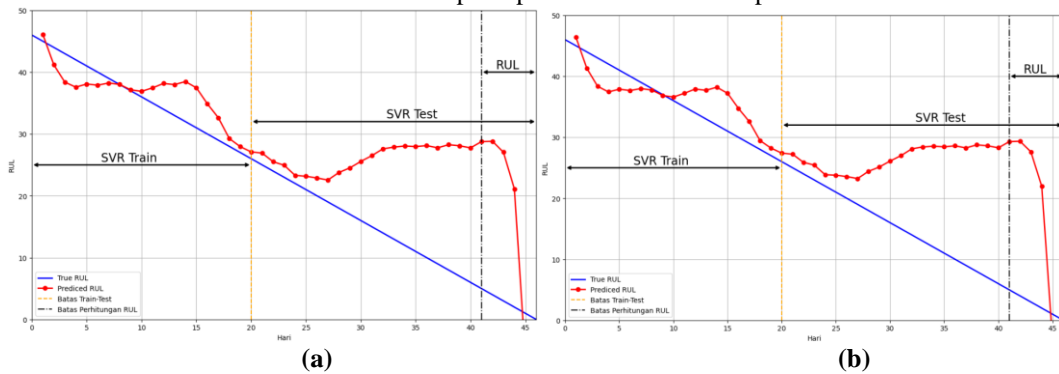
Prediksi RUL oleh model dengan *Gaussian smoothing* ditampilkan pada Gambar 6. Pada hasil prediksi, model kernel linear berhasil mencapai nilai RUL= 0 pada hari ke 43 sehingga model memprediksi bearing memiliki RUL selama 2 hari. Sama halnya dengan model kernel linear, model optuna pun memprediksi nilai RUL= 0 pada hari ke 43. Hasil keduanya memiliki kemiripan.



Gambar 6. Hasil Prediksi Nilai RUL dengan Input Enam Fitur Terbaik dengan Proses *Smoothing* Menggunakan Metode *Gaussian Smoothing* dengan Model (a) Kernel Linier (b) optuna

3.6.3. Model SVR *Exponential Moving Average*

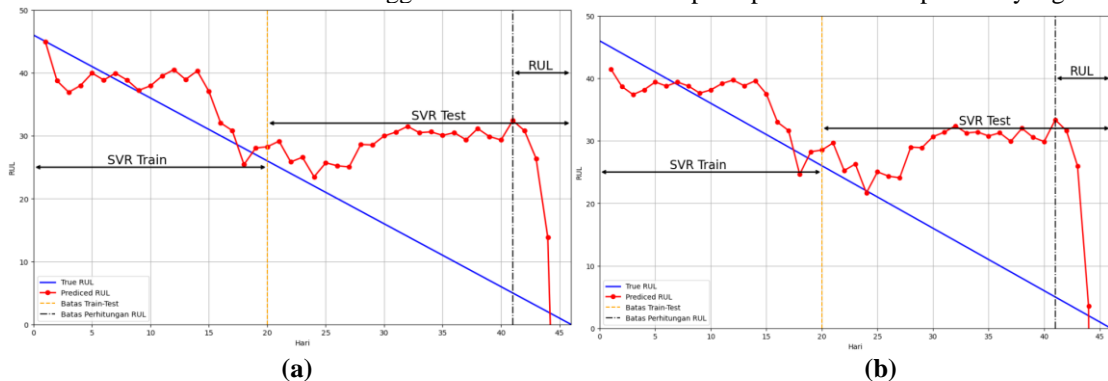
Prediksi RUL oleh model dengan input menggunakan proses *smoothing* metode *Exponential Moving Average* ditampilkan pada Gambar 7. Kedua model yang terbentuk mampu mendeteksi terjadinya kerusakan bantalan karena hasil prediksi yang mencapai nilai RUL=0 pada hari ke 45 sehingga model memprediksikan bearing memiliki RUL selama 4 hari. Dan hasil model kernel linear maupun optuna memiliki kemiripan.



Gambar 7. Hasil Prediksi Nilai RUL dengan Input Enam Fitur Terbaik dengan Proses *Smoothing* Menggunakan Metode *Exponential Moving Average* dengan Model (a) Kernel Linier (b) optuna

3.6.4. Model SVR Tanpa *Smoothing* dengan PCA

Gambar 8 menunjukkan prediksi RUL oleh model dengan input satu komponen PCA dari enam fitur terbaik tanpa proses *smoothing*. Untuk model dengan kernel linier dan parameter terbaik, hasil prediksi mencapai RUL = 0 pada hari ke 44. Sehingga model memiliki RUL selama 3 hari. Pada model optuna memprediksi RUL=0 pada hari ke 44. Sehingga model memiliki RUL selama 3 hari. Sehingga model kernel linear maupun optuna memiliki prediksi yang hampir mirip.

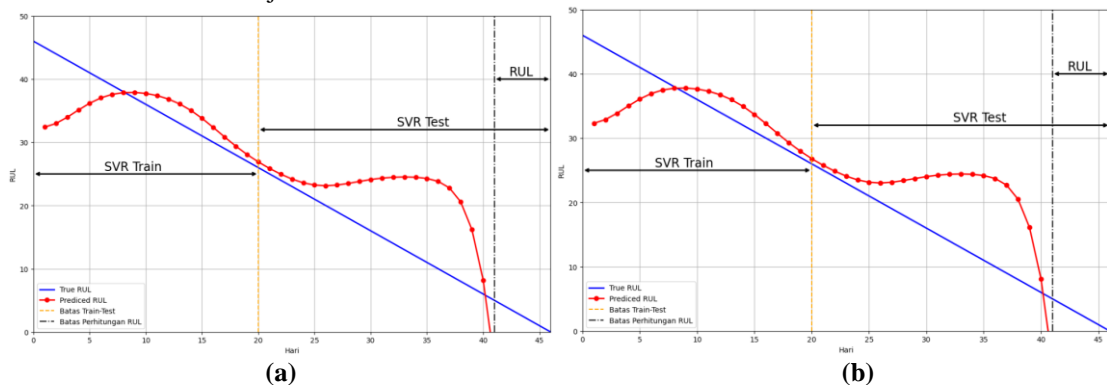


Gambar 8. Hasil Prediksi Nilai RUL dengan Input Satu Komponen PCA dari Enam Fitur Terbaik Tanpa Proses *Smoothing* dengan Model (a) Kernel Linier (b) optuna

3.6.5. Model SVR *Gaussian Smoothing* dengan PCA

Gambar 9 menunjukkan prediksi RUL *Gaussian Smoothing* yang mana kedua model yang dibentuk mampu mendeteksi terjadinya kegagalan bantalan karena prediksinya mencapai RUL=0. Namun, kedua model menghasilkan RUL yang diprediksi lebih rendah nilai dari nilai sebenarnya, yang berarti bahwa model meramalkan bahwa bantalan akan gagal lebih cepat dibandingkan kondisi aktual. Pada model kernel linier, terjadi *overfitting* karena terdapat perbedaan

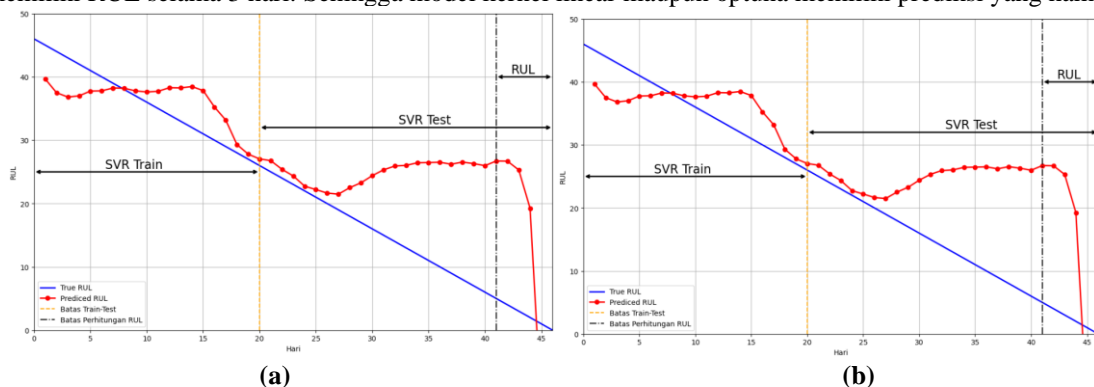
yang signifikan pada nilai RMSE *train* dan test. Dengan menggunakan input yang diambil dengan *proses smoothing*, prediksi model yang dihasilkan memiliki fluktuasi yang lebih kecil dibandingkan tanpa *smoothing*. Model dengan parameter terbaik memiliki kinerja lebih baik.



Gambar 9. Hasil Prediksi Nilai RUL dengan Input Satu Komponen PCA dari Enam Fitur Terbaik dengan Proses *Smoothing* Metode *Gaussian Smoothing* dengan Model (a) Kernel Linier (b) optuna

3.6.6. Model SVR *Exponential Moving Avarage* dengan PCA

Gambar 10 menunjukkan prediksi RUL oleh model dengan input satu komponen PCA dari enam fitur terbaik tanpa proses *smoothing*. Untuk model dengan kernel linier dan parameter terbaik, hasil prediksi mencapai RUL = 0 pada hari ke 44. Sehingga model memiliki RUL selama 3 hari. Pada model optuna memprediksi RUL=0 pada hari ke 44. Sehingga model memiliki RUL selama 3 hari. Sehingga model kernel linear maupun optuna memiliki prediksi yang hampir mirip.



Gambar 10. Hasil Prediksi Nilai RUL dengan Input Satu Komponen PCA dari Enam Fitur Terbaik dengan Proses *Smoothing* Metode *Exponential Moving Avarage* dengan Model (a) Kernel Linier (b) optuna

3.7. Perbandingan Variasi Performa Model SVR

Dari 12 variasi model SVR dengan beberapa input dan variasi parameter yang digunakan untuk membentuk model yang digunakan untuk memprediksi nilai RUL bearing, dapat dikelompokkan menjadi dua bagian berdasarkan hasil prediksi. Kelompok ini merupakan model yang dapat memprediksi kapan terjadinya kegagalan total pada bearing yaitu pada saat nilai RUL mencapai 0 dan model tidak dapat memprediksi kapan terjadinya kegagalan pada bearing karena dari hasil prediksi model tidak ada yang mencapai RUL sama dengan nol. Untuk model yang dapat memprediksi kapan terjadinya kegagalan total bantalan dirangkum pada Tabel 8 beserta nilai RMSE dan RUL-nya.

Tabel 8. Hasil Prediksi Variasi Model SVR

Variasi Model	RMSE Train	RMSE Test	RUL
Model SVR Linear tanpa proses <i>Smoothing</i>	3.758	39.268	3
Model SVR Optuna tanpa proses <i>Smoothing</i>	4,190	28.545	3
Model SVR PCA Linear tanpa proses <i>Smoothing</i>	5.956	21.139	5
Model SVR PCA Optuna tanpa proses <i>Smoothing</i>	5,549	21.139	5
Model SVR Linear dengan proses <i>Gaussian Smoothing</i>	1.118	15.295	2
Model SVR Optuna dengan proses <i>Gaussian Smoothing</i>	1.158	15.375	2
Model SVR PCA Linear dengan proses <i>Gaussian Smoothing</i>	5.332	34.783	0
Model SVR PCA Optuna dengan proses <i>Gaussian Smoothing</i>	5.375	34.811	0
Model SVR Linear dengan proses <i>Exponential Moving Avarage</i>	3.280	21.796	4

Model SVR Optuna dengan proses <i>Exponential Moving Avarage</i>	3.244	20.688	4
Model SVR PCA Linear dengan proses <i>Exponential Moving Avarage</i>	4.117	24.764	3
Model SVR PCA Optuna dengan proses <i>Exponential Moving Avarage</i>	4.128	24.633	3

Berdasarkan Tabel 8 dari 12 variasi model yang dibentuk menggunakan algoritma *Support Vector Regression* (SVR) dengan menggunakan beberapa konfigurasi parameter dan sejumlah variasi input dengan proses yang berbeda, 100% model yang dihasilkan dapat memprediksi sisa waktu (RUL) bearing sebelum mengalami kegagalan total yaitu pada saat RUL nol.

4. Kesimpulan

Kesimpulan dari penelitian ini adalah sebagai berikut:

1. Hasil dari penelitian ini dapat membuktikan bahwa pendekatan model SVR dapat digunakan untuk memprediksi *Remaining Useful Life* (RUL) dari NASA bearing dataset pada dataset ke-3.
2. Dalam melakukan prediksi *Remaining Useful Life* (RUL) dihasilkan 12 variasi model yang dapat dilihat pada tabel berikut.

Variasi Model	RMSE Train	RMSE Test	RUL
Model SVR Linear tanpa proses <i>Smoothing</i>	3.758	39.268	3
Model SVR Optuna tanpa proses <i>Smoothing</i>	4,190	28.545	3
Model SVR PCA Linear tanpa proses <i>Smoothing</i>	5.956	21.139	5
Model SVR PCA Optuna tanpa proses <i>Smoothing</i>	5,549	21.139	5
Model SVR Linear dengan proses <i>Gaussian Smoothing</i>	1.118	15.295	2
Model SVR Optuna dengan proses <i>Gaussian Smoothing</i>	1.158	15.375	2
Model SVR PCA Linear dengan proses <i>Gaussian Smoothing</i>	5.332	34.783	0
Model SVR PCA Optuna dengan proses <i>Gaussian Smoothing</i>	5.375	34.811	0
Model SVR Linear dengan proses <i>Exponential Moving Avarage</i>	3.280	21.796	4
Model SVR Optuna dengan proses <i>Exponential Moving Avarage</i>	3.244	20.688	4
Model SVR PCA Linear dengan proses <i>Exponential Moving Avarage</i>	4.117	24.764	3
Model SVR PCA Optuna dengan proses <i>Exponential Moving Avarage</i>	4.128	24.633	3

3. Dari berbagai metode SVR yang digunakan dalam melakukan analisis RUL NASA Bearing Dataset-3. Terpilih metode terbaik dalam kasus ini yaitu SVR Linear *Gaussian Smoothing* dengan dasar pemilihan adalah nilai RMSE mendekati angka 1

5. Daftar Pustaka

- [1] Gandhi. *Study of Vibration Generated in Steel Balls Creating Noise in Ball Bearings*. *International Research Journal of Engineering and Technology*. Open Journal of Applied Sciences. 2016;08(06):211–25.
- [2] Jin X, Que Z, Sun Y, Guo Y, Qiao W. *A Data-Driven Approach for Bearing Fault Prognostics*. *IEEE Trans Ind Appl*. 2019 Jul 1;55(4):3394–401.
- [3] Liu X, Song P, Yang C, Hao C, Peng W. *Prognostics and Health Management of Bearings Based on Logarithmic Linear Recursive Least-Squares and Recursive Maximum Likelihood Estimation*. *IEEE Transactions on Industrial Electronics*. 2017 Jul 29;65(2):1549–58.
- [4] Wang Y, Xu G, Liang L, Jiang K. *Detection of Weak Transient Signals Based on Wavelet Packet Transform and Manifold Learning for Rolling Element Bearing Fault Diagnosis*. *Mech Syst Signal Process*. 2015 Mar 1;54:259–76.
- [5] Abid K, Sayed-Mouchaweh M, Cornez L. *Adaptive Data-driven Approach for Fault Prognostics based on Normal Conditions-Application to Shaft Bearings of Wind Turbine*. 2019.
- [6] Wang D, Tsui KL, Miao Q. *Prognostics and Health Management: A Review of Vibration Based Bearing and Gear Health Indicators*. *IEEE Access*. 2017 Nov 19;6:665–76.
- [7] Heng A, Tan ACC, Mathew J, Montgomery N, Banjevic D, Jardine AKS. *Intelligent Condition-Based Prediction of Machinery Reliability*. *Mech Syst Signal Process*. 2009 Jul;23(5):1600–14.
- [8] Jianming S, Yongxiang L, Gong W, Xuzhi L. *Health Index Synthetization and Remaining Useful Life Estimation for Turbofan Engines Based on Run-To-Failure Datasets*. *Eksploatacja i Niezawodnosc*. 2016;18(4):621–31.
- [9] Yan W, Qiu H, Iyer N. *Feature Extraction for Bearing Prognostics and Health Management (PHM)-A Survey (Preprint) Stinfo Copy Air Force Research Laboratory Materials and Manufacturing Directorate Wright-Patterson Air Force Base, Oh 45433-7750 Air Force Materiel Command United States Air Force 1. Report Date (DD-MM-YY) 2. Report Type 3. Dates Covered (From-To) Feature Extraction For Bearing Prognostics And Health Management (Phm)-A Survey (Preprint)*. 2008;
- [10] Soualhi, A., Medjaher, K., & Zerhouni, N *Bearing Health Monitoring Based on Hilbert-Huang Transform, Support Vector Machine, and Regression*. *IEEE Transactions on Instrumentation and Measurement*. 2015 Jan;63(1):314–34.