

DIAGNOSA KETIDAKNORMALAN BERBASIS DATA DENGAN STUDI KASUS JET ENGINE

*Yoga Prasetya¹, Toni Prahasto², Achmad Widodo²

¹Mahasiswa Jurusan Teknik Mesin, Fakultas Teknik, Universitas Diponegoro

²Dosen Jurusan Teknik Mesin, Fakultas Teknik, Universitas Diponegoro

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

*E-mail: yogap131@gmail.com

Abstrak

Dalam era Industri 4.0, perusahaan semakin menggantungkan diri pada pemanfaatan data untuk mendapatkan wawasan yang berharga dalam menghadapi berbagai permasalahan. Salah satu tantangan krusial di sektor industri adalah memprediksi sisa umur dari jet engine, yang merupakan inti dari operasional penerbangan modern. Dengan meningkatnya kompleksitas dan jumlah data yang dihasilkan oleh sensor-sensor pada mesin, perusahaan kini dapat menerapkan analisis data dan teknik machine learning untuk meramalkan sisa umur jet engine dengan tingkat akurasi yang tinggi.

Penelitian ini menjelaskan bagaimana perusahaan mengadopsi pendekatan inovatif dengan memanfaatkan data sensor dan informasi historis untuk melatih model-machine learning. Proses ini melibatkan identifikasi pola, tren, dan anomali dalam data mesin guna membuat prediksi yang akurat terkait dengan kondisi dan performa jet engine. Dengan demikian, perusahaan dapat mengambil tindakan preventif secara tepat waktu, mengoptimalkan jadwal perawatan, dan mengurangi downtime yang tidak terduga.

Kata Kunci : analisis data; jet engine; machine learning; *remaining useful life*

Abstract

In the Industry 4.0 era, companies are increasingly relying on the use of data to gain valuable insights in dealing with various problems. One of the crucial challenges in the industrial sector is predicting the remaining life of jet engines, which are at the heart of modern aviation operations. With the increasing complexity and amount of data generated by sensors on engines, companies can now apply data analysis and machine learning techniques to predict the remaining life of jet engines with a high degree of accuracy.

This research explains how companies are adopting innovative approaches by leveraging sensor data and historical information to train machine-learning models. This process involves identifying patterns, trends and anomalies in engine data to make accurate predictions regarding jet engine condition and performance. This way, companies can take timely preventive action, optimize maintenance schedules, and reduce unexpected downtime.

Keywords : data analysis; jet engine; machine learning; *remaining useful life*

1. Pendahuluan

Dengan munculnya industri 4.0, terdapat ketersediaan data yang melimpah di area workshop manufaktur yang mengarahkan transisi menuju manufaktur yang mengandalkan ilmu pengetahuan, di mana strategi pemeliharaan inovatif seperti predictive maintenance dilakukan. Pada praktiknya, Predictive maintenance berhasil mengurangi biaya maintenance dan meminimalkan kemungkinan downtime produksi. Hal yang menjadi perhatian penting dalam predictive maintenance adalah prediksi *Remaining Useful Life* (RUL), yang berhubungan dengan sisa waktu komponen industri untuk menjalankan fungsi operasionalnya. Metode penilaian RUL dikelompokkan menjadi dua kategori utama: Pendekatan berbasis model dan berbasis data. Metode model fisik mengharuskan untuk memiliki pengetahuan terkait sistem dan mengeksplorasi sistem untuk membangun deskripsi matematis dari degradasinya sementara metode berbasis data (berbasis AI), melacak kondisi sistem dengan menggunakan *health indicator* (HI) yang diekstrak dari historis data sensor, untuk akhirnya memperkirakan RUL sistem. Karena tidak praktis untuk mengembangkan model matematika yang secara akurat menggambarkan fisika dan dinamika sistem manufaktur, metode berbasis data banyak menarik dan digunakan oleh akademisi dan industri [1,2].

Pada penelitian ini dilakukan analisis ketidaknormalan berbasis data dengan studi kasus *jet engine*. Data yang digunakan adalah dataset yang diukur oleh NASA dari pengukuran 21 unit sensor. Penelitian ini dilakukan dengan memodelkan dataset dengan tiga pemodelan algoritma yaitu KNN, SVM, dan Random Forest. Selanjutnya ketiga pemodelan ini dibandingkan untuk nilai performanya yaitu nilai R^2 dan RMSE.

2. Dasar Teori

2.1 Predictive maintenance

Predictive maintenance adalah pendekatan dalam manajemen pemeliharaan yang menggunakan analisis data dan teknik prediksi untuk mengidentifikasi kerusakan atau kegagalan pada peralatan atau mesin sebelum terjadinya kerusakan atau kegagalan, sehingga tindakan perbaikan atau pemeliharaan dapat dilakukan dengan tepat waktu [3].

Tujuan dari *predictive maintenance* adalah untuk mengoptimalkan kinerja peralatan, meminimalkan *downtime* (waktu tidak beroperasi), dan mengurangi biaya pemeliharaan.

2.2 Machine Learning

Machine learning secara luas mengacu pada proses menyesuaikan model prediktif ke data atau mengidentifikasi pengelompokan informatif dalam data. Melalui perhitungan, machine learning mencoba untuk mendekati atau meniru kemampuan manusia untuk mengenali pola, meskipun secara objektif. Misalnya, seorang anak yang belum pernah melihat anjing atau kucing tidak dapat membedakan mereka hanya dengan diberi informasi terperinci, tetapi mereka dapat belajar membedakan kedua hewan ini dengan melihat mereka berulang kali dan diberi tahu hewan mana yang mana. Dengan cara yang sama, ketika komputer diberikan sejumlah besar data dan algoritma untuk menganalisisnya, komputer dapat belajar melakukan tugas penilaian yang serupa dengan kemampuan belajar manusia. Ini disebut machine learning [4].

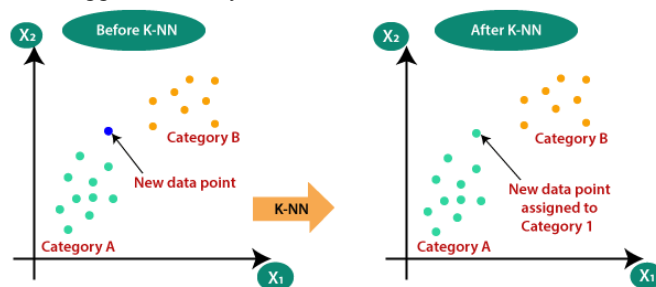
Secara umum, ada tiga jenis utama machine learning: supervised learning, unsupervised learning, dan reinforcement learning. Bentuk machine learning yang paling umum adalah supervised learning, di mana model dilatih dengan data berlabel, yaitu data masukan telah diberi label dengan label keluaran yang sesuai.

2.3 Pemodelan Algoritma Machine Learning

Pemodelan algoritma dalam konteks machine learning adalah proses merancang, melatih, dan mengevaluasi model yang dapat memahami pola dalam data dan membuat prediksi atau keputusan tanpa pemrograman eksplisit. Dalam konteks machine learning, "algoritma" sering kali merujuk pada model atau fungsi matematis yang dapat belajar dari data.

2.3.1 Pemodelan *K-Nearest Neighbour* (KNN)

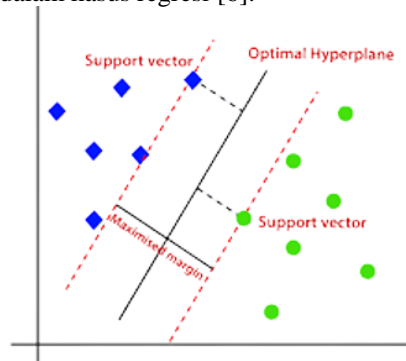
KNN (*K-Nearest Neighbors*) atau KNN adalah algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini bekerja dengan cara menentukan kelas atau nilai prediksi suatu data berdasarkan mayoritas dari kelas atau nilai data tetangga terdekatnya [5].



Gambar 2 Pemodelan KNN

2.3.2 Pemodelan *Super Vector Machine* (SVM)

SVM (Support Vector Machine) adalah algoritma machine learning yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini sangat efektif dalam menangani masalah klasifikasi dan regresi di ruang berdimensi tinggi. SVM bekerja dengan cara menemukan batas keputusan yang optimal atau *hyperplane* yang memisahkan dua kelas atau digunakan untuk memprediksi nilai target dalam kasus regresi [6].

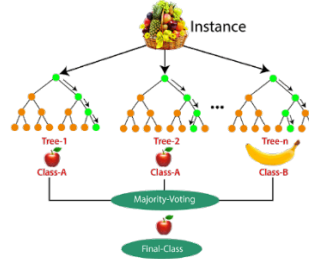


Gambar 3. Pemodelan SVM

2.3.3 Pemodelan Random Forest

Random Forest adalah sebuah algoritma dalam Machine Learning yang digunakan untuk melakukan tugas-tugas seperti klasifikasi, regresi, dan pemilihan fitur. Algoritma ini menggabungkan konsep dari metode Ensemble Learning dengan penggunaan *decision trees*.

Dalam klasifikasi suatu data, salah satu metode yang dapat digunakan adalah metode *Decision Tree* yang merupakan salah satu metode pengambilan keputusan. Pertumbuhan pada metode *Decision Tree* dapat ditumbuhkan sampai mencapai kehomogenan, tetapi hal tersebut dapat mengakibatkan *overfitting*, sehingga diperlukan suatu metode yang bisa mengatasi masalah tersebut. Metode yang dapat digunakan untuk menghindari *overfitting* adalah metode dengan memangkas pohon (*pruning*) atau menggunakan metode Random Forest (RF) [7].

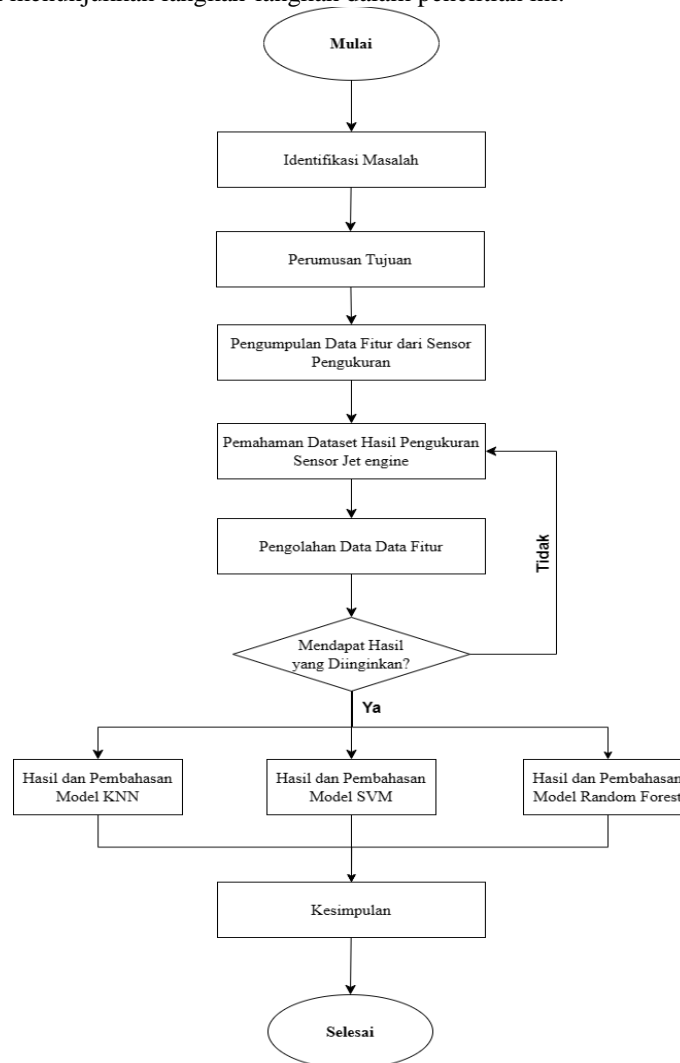


Gambar 4. Pemodelan Random Forest

3. Metode Penelitian

3.1 Diagram Alir Penelitian

Pada penelitian ini dilakukan beberapa langkah pengujian untuk melakukan analisis untuk memprediksi RUL *Jet Engine*. Gambar 5 dibawah menunjukkan langkah-langkah dalam penelitian ini.



Gambar 5. Diagram alir penelitian

3.2 Fitur-fitur *Jet Engine*

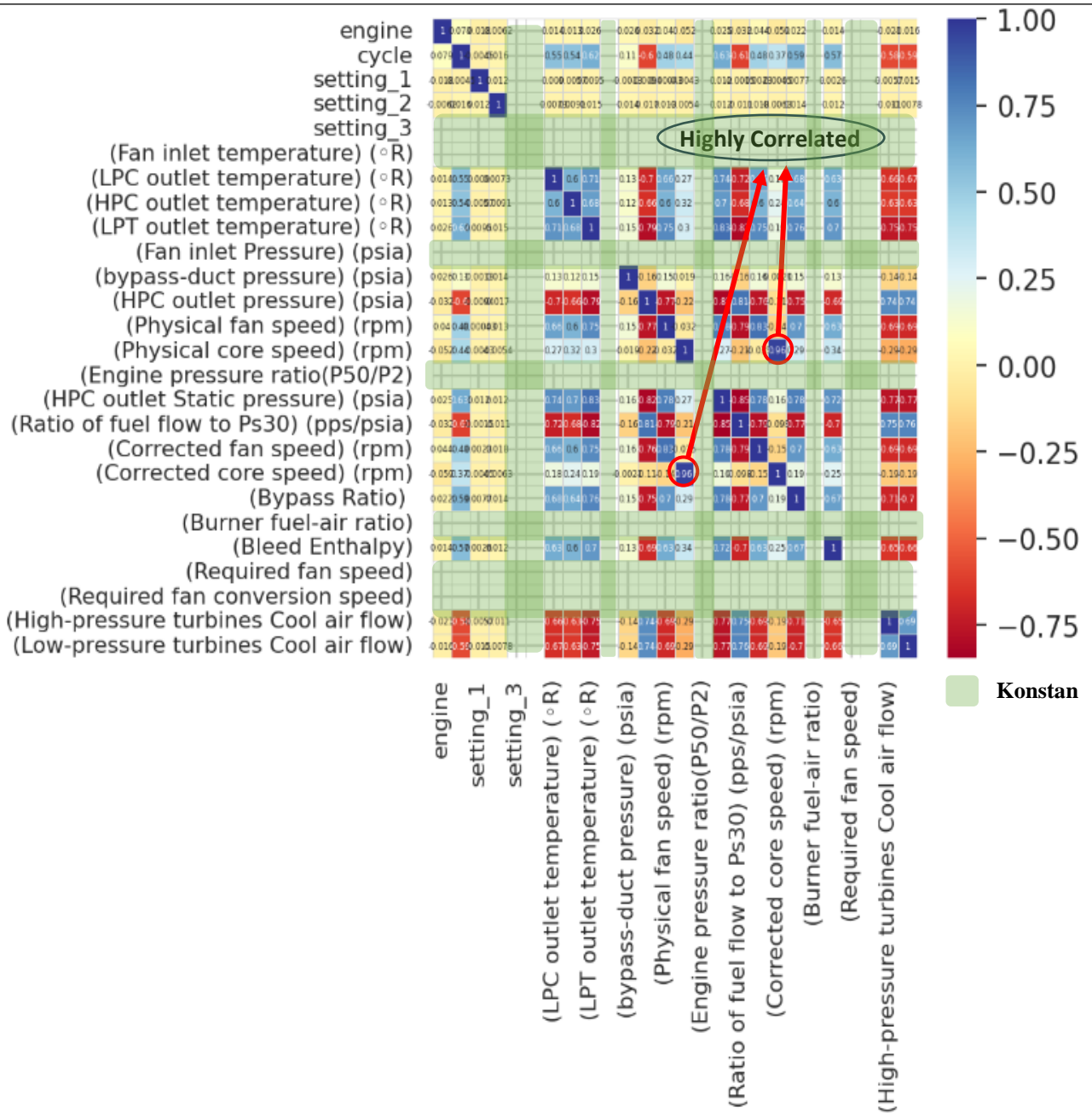
Fitur-fitur jet engine hasil pengukuran oleh sensor-sensor yang diekstrak oleh NASA memiliki 21 jumlah fitur. Fitur-fitur tersebut dapat dilihat pada Tabel 1 dibawah ini.

Tabel 1 Fitur-fitur *Jet Engine*

No. Sensor	Sensor Name	Unit
1.	Fan Inlet Temperature	°R
2.	Low Pressure Compressor (LPC) Outlet Temperature	°R
3.	High-Pressure Compressor (HPC) Outlet Temperature	°R
4.	Low-Pressure Turbine (LPT) Outlet Temperature	°R
5.	Fan Inlet Pressure	psia
6.	Bypass-Duct Pressure	psia
7.	HPC Outlet Pressure	psia
8.	Physical Fan Speed	rpm
9.	Physical Core Speed	rpm
10.	Engine Pressure Ratio (P50/P2)	-
11.	HPC Outlet Static Pressure	psia
12.	Ratio Of Fuel Flow To Ps30	pps/psia
13.	Corrected Fan Speed	rpm
14.	Corrected Core Speed	rpm
15.	Bypass Ratio	-
16.	Burner Fuel-Air Ratio	-
17.	Bleed Enthalpy	-
18.	Required Fan Speed	-
19.	Required Fan Conversion Speed	-
20.	High-Pressure Turbines Cool Air Flow	-
21.	Low-Pressure Turbines Cool Air Flow	-

3.3 Korelasi Data antar Fitur

Korelasi data adalah ukuran yang digunakan untuk menilai hubungan statistik antara dua atau lebih variabel dalam sebuah dataset. Ini membantu kita memahami sejauh mana variabel-variabel tersebut berkaitan satu sama lain. Korelasi dapat membantu mengidentifikasi apakah ada hubungan positif, negatif, atau tidak ada hubungan sama sekali antara variabel-variabel tersebut. Pada Gambar 6 dibawah ini adalah korelasi data antar fitur-fitur *jet engine* [8, 9].



Gambar 6. Korelasi Data antar Fitur-fitur Jet Engine

Pada Gambar 6 diatas terdapat beberapa fitur yang berkorelasi sangat tinggi sehingga akan mengakibatkan data yang berulang jika dilibatkan dalam pemodelan algoritma, sehingga salah satu dari fitur-fitur tersebut harus dihilangkan. Fitur-fitur tersebut adalah *Physical Core Speed & Corrected Core Speed*. Selain itu juga terdapat fitur-fitur yang bernilai konstan yaitu *Setting 3, Fan inlet Temperature (°R), Fan inlet Pressure (psia), Engine pressure ratio (P50/P2), Burner Fuel-Air Ratio, Required Fan Speed, dan Required Fan Conversion Speed*. Sehingga untuk fitur-fitur yang berkorelasi tinggi dan bernilai konstan tersebut dapat dilihat pada Tabel 2 dibawah ini.

Tabel 2 Fitur-fitur yang berkorelasi tinggi dan bernilai konstan

Fitur Berkorelasi Tinggi	Fitur Bernilai Konstan
1. Physical Core Speed	1. Setting 3
2. Corrected Core Speed	2. Fan inlet Temperature (°R)
	3. Fan inlet Pressure (psia)
	4. Engine pressure ratio (P50/P2)
	5. Burner Fuel-Air Ratio

	6. Required Fan Speed
	7. Required Fan Conversion Speed

Setelah fitur-fitur tersebut teridentifikasi, selanjutnya fitur-fitur pada Tabel 2 di atas dieliminasi untuk membuat pemodelan yang berkinerja tinggi. Sehingga fitur-fitur yang tersisa untuk pemodelan dapat dilihat pada Tabel 3 sebagai berikut.

Tabel 3 Fitur-fitur Pemodelan

Fitur-fitur Pemodelan	
1. Engine	10. HPC outlet Static pressure (psia)
2. Cycle	11. Physical Fan Speed (rpm)
3. Setting 1	12. Physical Core Speed (rpm)
4. Setting 2	13. Ratio of fuel flow to Ps30 (pps/psia)
5. LPC Outlet Temperature (°R)	14. Corrected Fan Speed (rpm)
6. HPC Outlet Temperature (°R)	15. Bypass Ratio
7. LPT Outlet Temperature (°R)	16. Bleed Enthalpy
8. Bypass-Duct Pressure	17. High-pressure turbines Cool air flow
9. HPC outlet pressure (psia)	18. Low-pressure turbines Cool air flow

3.4 Seleksi Fitur

Backward regression adalah salah satu metode seleksi fitur yang dimulai dengan semua fitur dan secara berulang menghapus fitur satu per satu berdasarkan kriteria tertentu hingga model yang dihasilkan dianggap memadai. Metode ini biasanya digunakan dalam konteks regresi, di mana tujuan utamanya adalah untuk meminimalkan *overfitting* dan meningkatkan interpretabilitas model.

Pada seleksi fitur menggunakan *backward regression*, diperlukan pengukuran signifikansi fitur-fitur jet engine dengan mengidentifikasi P-Value dari setiap fitur jet engine. Untuk P-Value atau $P < 0,05$ dikategorikan signifikan untuk pemodelan prediksi target variabel. Sedangkan untuk $P > 0,05$ tidak signifikan untuk pemodelan prediksi target variabel. Berikut adalah hasil P-Value untuk setiap fitur *jet engine*.

Tabel 4 Fitur-fitur Jet Engine

No.	Fitur-fitur Jet Engine	P-Value
1.	Cycle	0.000
2.	Setting_1	0.455
3.	Setting_2	0.062
4.	LPC Outlet Temperature (°R)	0.000
5.	HPC Outlet Temperature (°R)	0.000
6.	LPT Outlet Temperature (°R)	0.000
7.	Bypass-duct pressure (psia)	0.000
8.	HPC outlet pressure (psia)	0.000
9.	Physical fan speed (rpm)	0.000
10.	Physical core speed (rpm)	0.000
11.	HPC outlet Static pressure (psia)	0.000
12.	Ratio of fuel flow to Ps30 (pps/psia)	0.000
13.	Corrected fan speed (rpm)	0.000
14.	Bypass Ratio	0.000
15.	Bleed Enthalpy	0.000
16.	High-pressure turbines Cool air flow	0.000
17.	Low-pressure turbines Cool air flow	0.000

Dari Tabel 4 di atas terdapat dua fitur yang tidak signifikan yaitu *Setting 1* dan *Setting 2*, sehingga tidak perlu dimasukkan kedalam pemodelan prediksi RUL. Jadi untuk fitur-fitur yang tersisa setelah seleksi fitur dapat dilihat pada Tabel 5 dibawah ini.

Tabel 5 Fitur hasil seleksi dengan *backward regression*

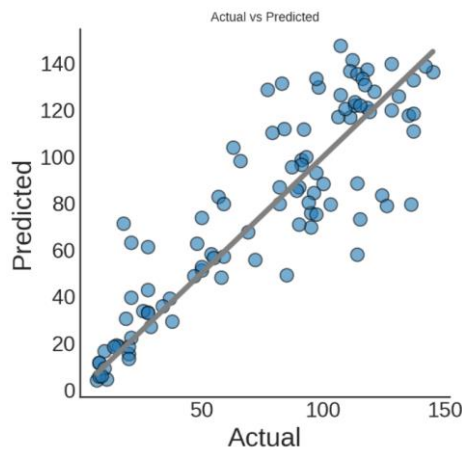
No	Fitur-fitur Jet Engine Setelah Seleksi
1.	Cycle
2.	LPC Outlet Temperature (°R)

- | | |
|-----|---------------------------------------|
| 3. | HPC Outlet Temperature (°R) |
| 4. | LPT Outlet Temperature (°R) |
| 5. | Bypass-duct pressure (psia) |
| 6. | HPC outlet pressure (psia) |
| 7. | Physical fan speed (rpm) |
| 8. | Physical core speed (rpm) |
| 9. | HPC outlet Static pressure (psia) |
| 10. | Ratio of fuel flow to Ps30 (pps/psia) |
| 11. | Corrected fan speed (rpm) |
| 12. | Bypass Ratio |
| 13. | Bleed Enthalpy |
| 14. | High-pressure turbines Cool air flow |
| 15. | Low-pressure turbines Cool air flow |

4. Hasil dan Pembahasan

4.1 Pemodelan K-Nearest Neighbour (KNN)

Dengan menggunakan parameter nilai K = 9, maka didapat hasil pemetaan nilai RUL aktual dan nilai RUL terprediksi untuk pemodelan KNN, yang dapat dilihat pada Gambar 7 dibawah ini.



Gambar 7 Grafik *scatterplot* RUL aktual dan RUL terprediksi

Pada pemodelan KNN ini didapat nilai R^2 dan RMSE adalah sebagai berikut.

Tabel 6 Nilai akurasi pemodelan KNN

	R^2	RMSE
KNN	0,774524	19,732389

4.2 Pemodelan Super Vector Machine (SVM)

Dengan menggunakan parameter berikut ini pada pemodelan Super Vector Machine (SVM),

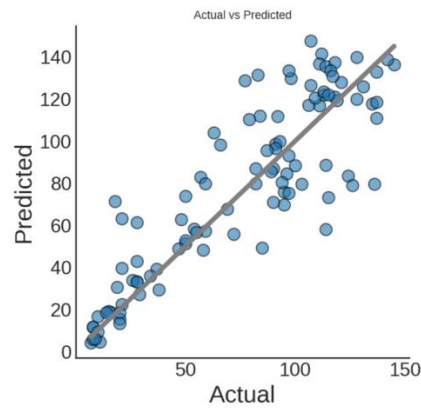
kernel = rbf,

C = 100,

Gamma = 0.5,

Epsilon = 0.01

maka didapat hasil pemetaan nilai RUL aktual dan nilai RUL terprediksi untuk pemodelan SVM, yang dapat dilihat pada Gambar 8 dibawah ini.



Gambar 8 Pemodelan *Super Vector Machine* (SVM)

Pada pemodelan SVM ini didapat nilai R^2 dan RMSE adalah sebagai berikut.

Tabel 7 Nilai akurasi pemodelan SVM

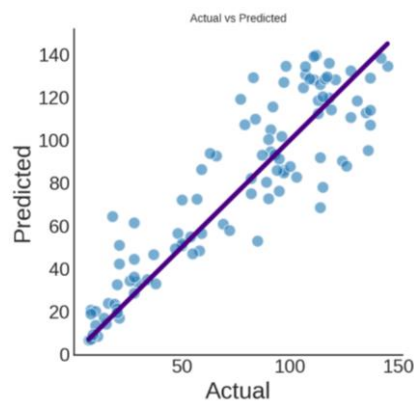
	R^2	RMSE
SVM	0,737049	21,309221

4.3 Pemodelan Random Forest

Dengan menggunakan parameter berikut ini pada pemodelan random forest,

- N jobs = -1
- N Estimator = 500
- Min. Sample Leaf = 1
- Max. Features = sqrt

maka didapat hasil pemetaan nilai RUL aktual dan nilai RUL terprediksi untuk pemodelan *random forest*, yang dapat dilihat pada Gambar 9 dibawah ini.



Gambar 9 Pemodelan *Random Forest*

Pada pemodelan *random forest* ini didapat nilai R^2 dan RMSE adalah sebagai berikut.

Tabel 8 Nilai akurasi pemodelan *random forest*

	R ²	RMSE
Random Forest	0,789700	19,056793

5. Kesimpulan

Dari pembahasan dan analisis yang telah dilakukan dalam penelitian, maka dapat disimpulkan bahwa:

- Fitur-fitur yang berpengaruh untuk memprediksi nilai RUL (Target Variabel) adalah berjumlah 15 Fitur yaitu:
 - Cycle
 - LPC Outlet Temperature (°R)
 - HPC Outlet Temperature (°R)
 - LPT Outlet Temperature (°R)
 - Bypass-duct pressure (psia)
 - HPC outlet pressure (psia)
 - Physical fan speed (rpm)
 - Physical core speed (rpm)
 - HPC outlet Static pressure (psia)
 - Ratio of fuel flow to Ps30 (pps/psia)
 - Corrected fan speed (rpm)
 - Bypass Ratio
 - Bleed Enthalpy
 - High-pressure turbines Cool air flow
 - Low-pressure turbines Cool air flow
- Pemodelan KNN, SVM, dan *Random Forest* digunakan untuk membuat model prediksi RUL Jet Engine dengan keakuratan yang dapat dilihat pada Tabel 9 berikut:

Tabel 9 Nilai akurasi pemodelan KNN, SVM, dan *Random Forest*

	R ²	RMSE
KNN	0,774524	19,732389
SVM	0,737049	21,309221
Random Forest	0,789700	19,056793

Untuk pemodelan dengan kinerja terbaik adalah pemodelan dengan nilai R² tertinggi dan nilai RMSE terendah [10] yaitu pemodelan *Random Forest*.

DAFTAR PUSTAKA

- Marios Tirovolas, a. C. S., 2022. Introducing Fuzzy Cognitive Map for Predicting Engine's Health Status. *ScienceDirect*.
- Nazara, K. Y., 2022. Perancangan Smart Predictive Maintenance untuk Mesin Produksi. Seminar Nasional Official Statistics, pp. 691-695.
- Wahid, E. A., 2022. Pengembangan Algoritma Predictive Maintenance Pada Coal Pfister Feeder Dengan Pendekatan Machine Learning. *Sains, Aplikasi, Komputasi dan Teknologi Informasi*, Volume 4, p. 22.
- Jin Komuro, D. K. H. H. S. Y., 2023. Machine learning in cardiology: Clinical application and basic research. *Journal of Cardiology*.
- A. Ali, M. H. N. G. e. a., 2023. A k nearest neighbour ensemble via extended neighbourhood rule and feature subsets. *Elsevier*.
- Fatma Gul Altin, I. B. a. F. O., 2023. Predicting the Amount of Medical Waste Using Kernel Based SVM and Deep Learning Methods For A Private Hospital in Turkey. *Sustainable Chemistry and Pharmacy*.
- Amaliah, S., 2022. Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng. *Journal of Statistics and Its Application on Teaching and Research*, Volume 4, p. 121.
- Wibowo, R. A., 2019. Analisis Korelasi Dalam Penentuan Arah Antar Faktor Pada Pelayanan Angkutan Umum Di Kota Magelang. *Journal of Electrical Engineering, Computer and Information Technology*.

-
- [9] Suparto, 2014. Analisis Korelasi Variabel - Variabel Yang Mempengaruhi siswa Dalam Memilih Perguruan Tinggi. IPTEK, Volume 18.
- [10] Imam Suprayogi, T., 2012. Model Prediksi Liku Kalibrasi Menggunakan Pendekatan Jaringan Saraf Tiruan (JST).