

Predicting Machine Failure Type

Zefanya Frandita Situmorang || 121320045 || Sains Data Rekayasa || Teknik Fisika
Dosen Pengampu : Ahmad Suaif, S.Si., M.Si.

1 PROJECT DOMAIN

Prediksi kerusakan pada alat produksi penting dalam menjaga kelancaran produksi. Predicting Machine Failure Type adalah metode yang memonitor status mesin untuk mencegah kegagalan. Memanfaatkan pengumpulan data dan analisis Machine Learning (ML) untuk memprediksi kegagalan peralatan, khususnya menggunakan metode regresi untuk memodelkan hubungan antar variabel.

2 BUSSINESS UNDERSTANDING

Predicting Machine Failure Type adalah sistem yang memiliki potensi signifikan sebagai alat bantu dalam pengambilan keputusan pada bidang industri dan manufaktur serta dapat meningkatkan efisiensi operasional yang membantu perusahaan tetap kompetitif di pasar dinamis.

3 DATA UNDERSTANDING

Sumber Dataset: Machine Predictive Maintenance Classification.

URL:

<https://www.kaggle.com/datasets/shivamb/machine-predictive-maintenance-classification>

Detail Dataset:

- Dataset berupa CSV
- Terdiri dari 10000 records, 10 buah fitur yang diukur.
- Terdiri dari 3 data kategori, dan 7 data numerik.

Variabel dataset:

- UID
- Product ID
- Type
- Air Temperature [K]
- Process Temperature [K]
- Rotational Speed [rpm]
- Torque [Nm]
- Tool Wear [min]
- Target
- Failure Type:
 - Tool Wear Failure (TWF)
 - Heat Dissipation Failure (HDF)
 - Power Failure (PWF)
 - Overstrain Failure (OSF)
 - Random Failures (RNF)

4 DATA PREPARATION

Dilakukan kegiatan Data Gathering, Data Assessing, dan Data Cleaning. Pada Data Cleaning, dilakukan penanganan terhadap Anomalie Target dengan melakukan dropping fitur 'Random Failures' serta Outliers. Pada Pre-Processing, dilakukan Encoding, Reduction dengan PCA dan Oversampling with SMOTE.

Berdasarkan boxplot Gambar 4.1, tidak dilakukan penanganan terhadap kemungkinan outlier. Hal ini dikarenakan dapat menghilangkan informasi yang diperlukan.

ENCODING

Type	Encoded	Count
L	0	5984
M	1	2991
H	2	998

Tabel 4.1 Encoding Fitur 'Type'

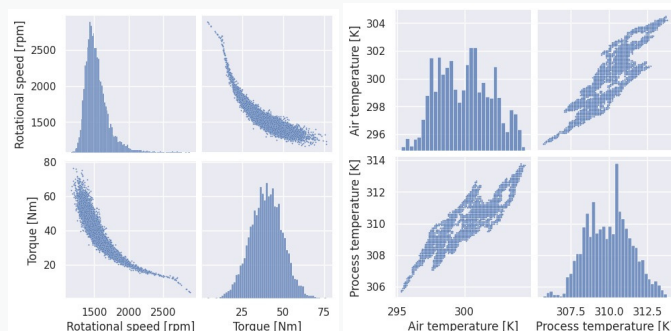
Type/Failure Type	Heat Dissipation Failure	No Failure	Overstrain Failure	Power Failure	Random Failures	Tool Wear Failure
H	8	979	1	5	4	6
L	74	5757	73	59	12	25
M	30	2916	4	31	2	14

Tabel 4.2 Encoding Fitur 'Failure Type'

Encoding data adalah proses mengonversi variabel kategoris menjadi bentuk numerik sehingga dapat dimengerti oleh algoritma machine learning atau model statistik.

REDUCTION with PCA

Reduksi dimensi dengan PCA digunakan untuk mengurangi jumlah fitur (variabel) dalam dataset, dengan tetap mempertahankan sebanyak mungkin informasi yang mungkin.

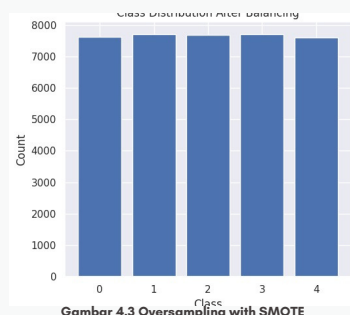


Gambar 4.2 Reduction with PCA

OVERSAMPLING with SMOTE

Setelah data dibersihkan, dataset dibagi menjadi data train dan data test dengan rasio pembagian 80:20. Dikarenakan jumlah kategori pada fitur 'Failure Type' tidak seimbang, dilakukan Oversampling dengan SMOTE. Berikut adalah detail dari dataset:

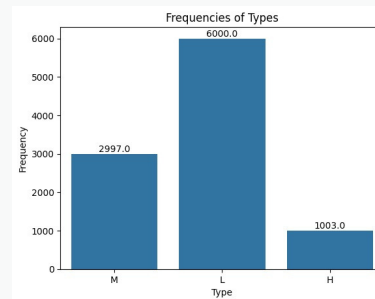
- Total sampel dataset train : 38306
- Total sampel dataset test : 1995



Gambar 4.3 Oversampling with SMOTE

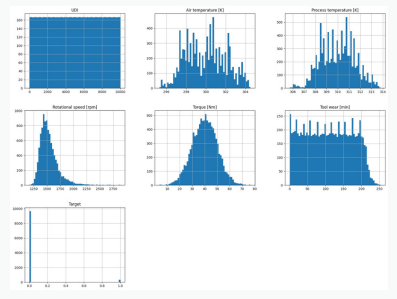
3.1 DATA ANALYSIS EDA

ANALISIS UNIVARIAT



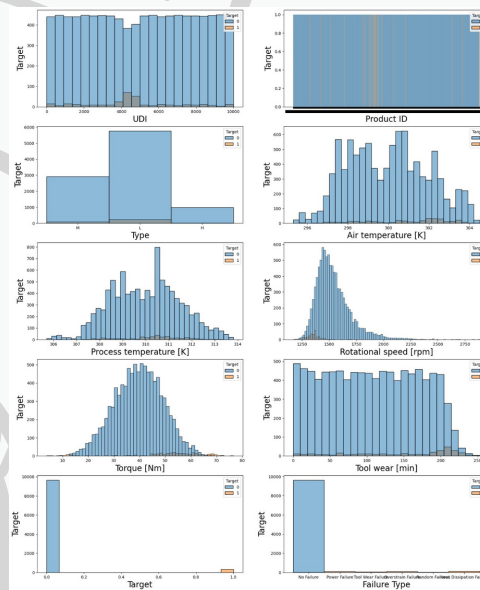
Gambar 3.1 Analisis Univariate Fitur 'Type' (Data Kategori)

Gambar 3.2 Analisis Univariate Fitur 'Failure Type' (Data Kategori)

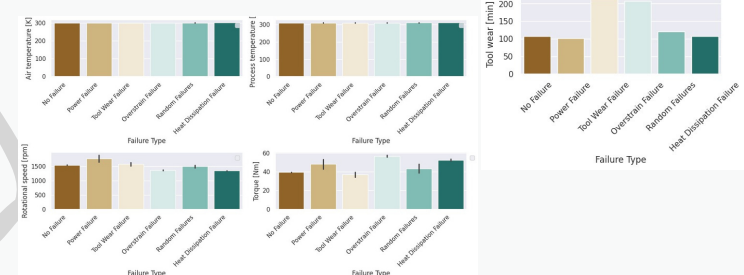


Gambar 3.3 Analisis Univariate Data Numerik

ANALISIS BIVARIAT



Gambar 3.4 Analisis Bivariate antara fitur kategori 'Target' dengan Fitur dataset

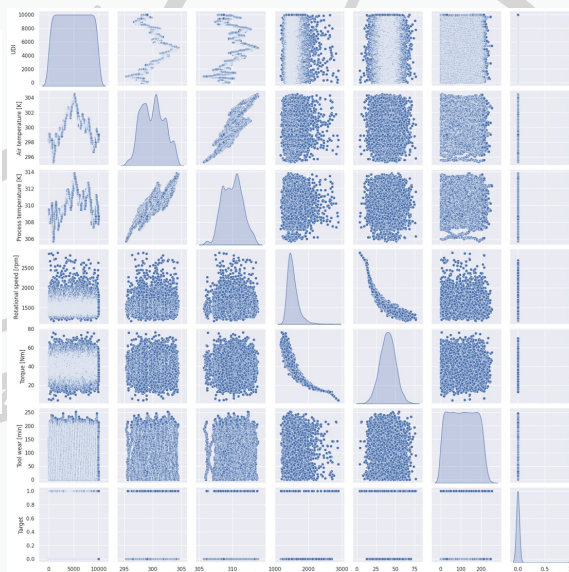


Gambar 3.5 Analisis Bivariate antara Data Numerik dengan Failure Type

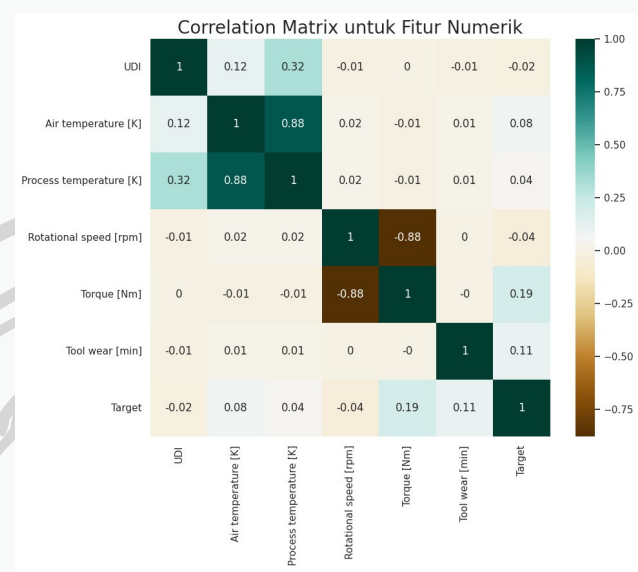
Type/Failure Type	Heat Dissipation Failure	No Failure	Overstrain Failure	Power Failure	Random Failures	Tool Wear Failure
H	8	979	1	5	4	6
L	74	5757	73	59	12	25
M	30	2916	4	31	2	14

Tabel 3.1 Analisis Bivariate fitur 'Type' dengan 'Failure Type'

ANALISIS MULTIVARIAT



Gambar 3.7 Analisis Multivariat



Gambar 3.8 Correlation Matrix Numeric Fitur

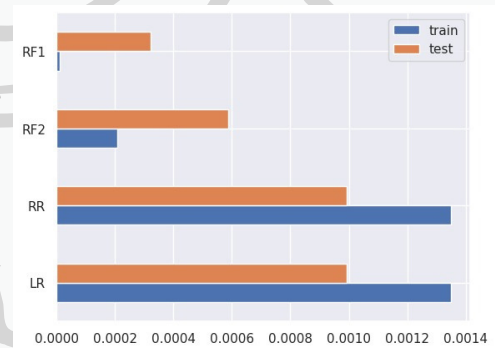
5 MODEL

$$y = a + bX$$

- y adalah variabel kriterium.
- a adalah intersep.
- b adalah slope.
- X adalah variabel prediktor.

Pada proyek Predictive Machine Failure System, digunakan 4 macam algoritma regresi:

- Linear Regression
- Ridge Regression
- Random Forest Regresor
- Random Forest Regresor - Tuning GridSearchCV



Gambar 5.5 Perbandingan Model berdasarkan Nilai Error

6 EVALUATION

	Model 1	Model 2	Model 3	Model 4
R2	-14.257.122.047.889.900	-14.256.282.814.195.700	-7.092.059.116.469.190	-923.334.987.107.827
MSE	11.751.993.191.251.800	11.751.647.302.909.400	0.8853675447084435	0.9956414891470533
MAE	8.315.954.515.654.350	8.315.714.189.178.560	0.6505664160401003	0.7491044047744142

Tabel 5.1 Perbandingan Performa MAE, MSE, dan R2 Model

Model 3 (RF1) dan Model 4 (RF2) menampilkan hasil performa yang lebih baik. Nilai error train dan test dari Model 3 (RF1) dan Model 4 (RF2) jauh lebih baik dibandingkan model lainnya.

	y_true	prediksi_LR	prediksi_RR	prediksi_RF1	prediksi_RF2
4979	1	0.0	0.0	1.0	1.0

Tabel 5.2 Perbandingan Model

Model yang telah dikembangkan dapat memprediksi tipe kerusakan mesin dengan baik dengan menggunakan **Random Forest Regressor**.