

DATA SCIENTIST PROJECT

DATA UNDERSTANDING

Dataset yang digunakan merupakan data pengukuran nilai vibrasi / getaran diukur dari bearing poros berkecepatan tinggi di mana data ini disediakan oleh Green Power Monitoring System (GPMS). Pengukuran nilai vibrasi diambil selama 50 hari dan dihari akhir terdapat kerusakan pada bagian inner race.

Data vibrasi pada high speed shaft (HSS) bearing didapatkan dengan cara mengukur nilai vibrasi suatu HSS bearing yang terpasang pada turbin angin 2 MW menggunakan akselerometer berbasis MEMS (Micro - Electromechanical System) dan terpasang secara radial di cincin bearing. Pengukuran ini menggunakan satuan percepatan. Data vibrasi direkam setiap harinya selama 6 detik dengan laju pengambilan sampel data sebesar 97,656 Hz. Oleh karena itu, setiap harinya data yang berhasil terekam sebanyak 585.936 sampel data dan jika ditotalkan selama 50 hari, maka total data yang berhasil terekam sebanyak 29.296.800 data.

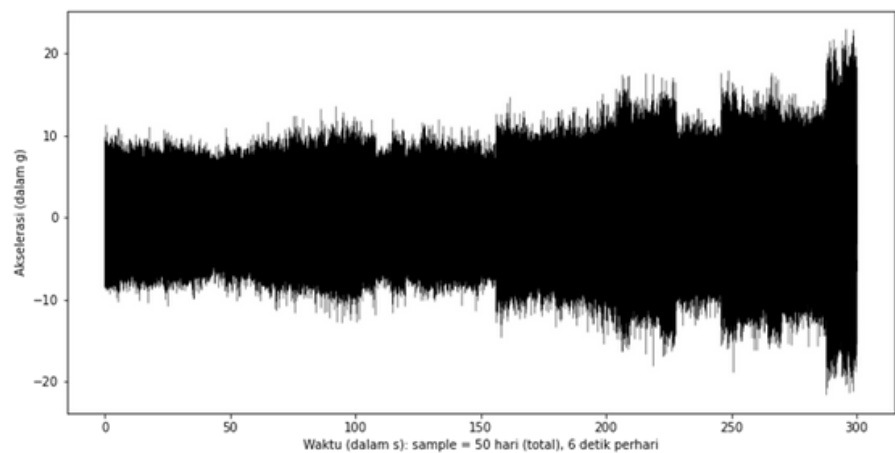
Data vibrasi kemudian diekstrak fitur time domainnya meliputi mean absolut, standar deviasi, skewness, kurtosis, entropi, rms, maksimum absolut, peak to peak, crest factor, faktor clearance, faktor shape, dan impuls. fitur time domain tersebut kemudian dianalisis dan digunakan untuk membangun model machine learning.

METHODS

- Segmentasi data
- Ekstraksi fitur time domain
- data cleansing
- seleksi fitur
- normalisasi data
- train test split
- Model training (Gradient boosting algorithm)
- Evaluation
- Deployment

DATA SCIENTIST PROJECT

RESULT

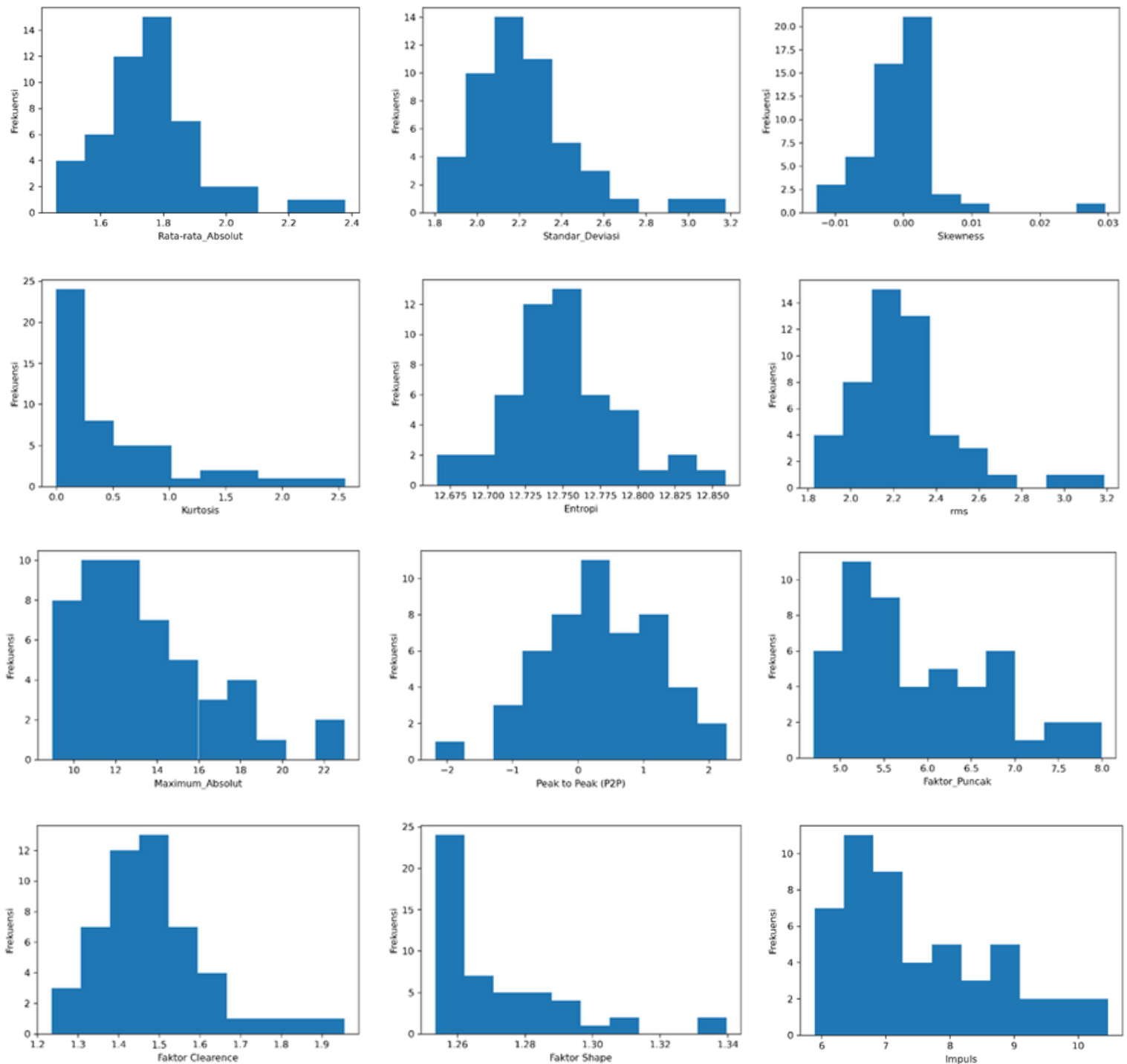


Sinyal vibrasi secara keseluruhan dengan total waktu rekam selama 300 detik. Terlihat bahwa nilai kecepatan vibrasi bearing relatif terus meningkat yang mengindikasikan adanya kondisi tidak wajar pada bearing.

	Rata-rata_Absolut	Standar_Deviasi	Skewness	Kurtosis	Entropi	Root_Mean_Square_(RMS)	Maximum_Absolut	Peak_to_Peak_(P2P)	Faktor_Puncak	Faktor_CI
0	1.832206	2.270483	0.003870	-0.004378	12.768751	2.296702	11.287679	0.954289	4.914735	1
1	1.655326	2.062148	0.003010	0.019467	12.726319	2.076542	10.201767	1.093627	4.912863	1
2	1.686617	2.103555	-0.001029	0.022406	12.736041	2.114894	11.027756	0.581436	5.214330	1
3	1.608343	2.008065	0.001477	0.041467	12.713978	2.019404	10.674653	1.829304	5.286041	1
4	1.650317	2.060564	0.001012	0.044484	12.726632	2.071768	10.846172	-0.475386	5.235226	1
5	1.674787	2.079129	-0.008428	0.018007	12.730569	2.099721	10.071461	0.093068	4.796572	1
6	1.581854	1.971982	-0.001429	0.017419	12.707029	1.983443	9.619002	0.401022	4.849649	1
7	1.457977	1.811430	0.002216	0.005757	12.666038	1.827789	8.971681	-0.081710	4.908489	1
8	1.597980	1.997333	0.001559	0.071093	12.710461	2.008323	10.909793	0.699154	5.432289	1
9	1.593219	1.984205	-0.001959	0.007185	12.709072	1.997857	10.085749	1.339522	5.048283	1
10	1.687565	2.113043	-0.001371	0.124720	12.736072	2.124111	11.652299	1.446902	5.485731	1
11	1.703601	2.133507	-0.012744	0.093440	12.739059	2.142290	11.396905	-1.205131	5.319964	1
12	1.698811	2.128358	-0.000204	0.164703	12.738538	2.139613	12.384752	0.718744	5.788313	1
13	1.815225	2.253587	-0.002308	0.081683	12.763997	2.281946	12.042024	1.450914	5.277085	1
14	1.735302	2.178178	-0.000193	0.154839	12.748158	2.186517	13.233013	0.951470	6.052097	1
15	1.738139	2.186108	0.002037	0.269095	12.747229	2.195762	13.559437	0.679500	6.175276	1
16	1.743279	2.198306	0.000509	0.347857	12.748150	2.208262	12.791482	-0.493855	5.792556	1
17	1.777740	2.209385	0.001703	0.109138	12.755961	2.236438	11.938507	-0.010251	5.338178	1
18	1.543785	1.927763	0.002140	0.034676	12.696112	1.937719	9.095206	0.372127	4.693771	1
19	1.784250	2.228535	-0.001033	0.092083	12.759001	2.243593	12.037040	0.362480	5.365073	1
20	1.548290	1.923362	-0.007469	0.137991	12.692676	1.949847	10.182065	0.425804	5.221983	1
21	1.794219	2.242475	0.001566	0.131236	12.760108	2.259990	11.579474	1.323277	5.123684	1

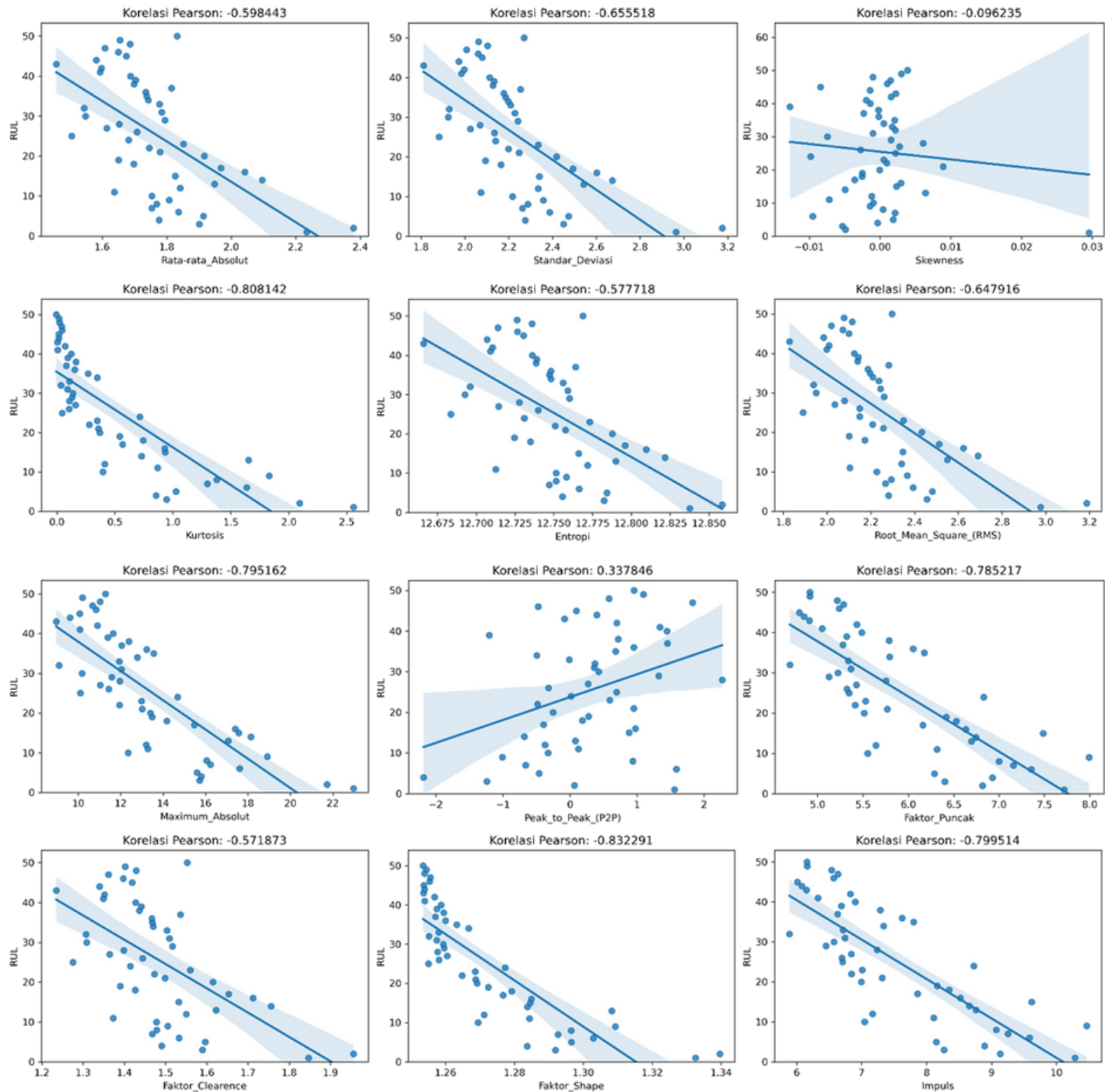
Tampilan dataframe dari fitur time domain sinyal vibrasi bearing yang telah diekstraksi dengan dimensinya adalah 12 kolom dan 50 baris.

DATA SCIENTIST PROJECT



Gambar di atas merupakan kurva distribusi dari masing-masing fitur time domain sinyal vibrasi. dari kurva tersebut, fitur didominasi oleh distribusi miring ke kanan karena data lebih banyak terkumpul di sekitar nilai yang lebih rendah. Terdapat data outlier yang dalam konteks analisis ini, data outlier tersebut adalah nilai pada akhir waktu rekam yang menandakan data outlier tersebut adalah pertanda adanya ketidakwajaran pada bearing atau bearing mengalami degradasi.

DATA SCIENTIST PROJECT



Gambar di atas adalah nilai koefisien fitur vibrasi terhadap nilai RUL serta visualisasinya untuk menggambarkan seberapa linearkah masing-masing fitur time domain terhadap nilai RUL. Nilai koefisien korelasi beragam dari nilai positif dan nilai negatif. Nilai positif menunjukkan hubungan searahantara dua fitur, sedangkan nilai negatif menunjukkan hubungan terbalik antara dua fitur. Nilai koefisien korelasi diabsolutkan untuk kemudian diambil fitur dengan nilai koefisien korelasi yang kuat dan sangat kuat. Dalam statistika, nilai koefisien korelasi dengan interpretasi kuat berada pada range 0.60 – 0.799. Sedangkan nilai koefisien korelasi dengan interpretasi sangat kuat berada pada range 0.8 – 1.0. Oleh karena itu, fitur time domain yang terseleksi untuk dipakai pada tahap membangun model adalah fitur yang memiliki koefisien korelasi antara 0.6 sampai dengan 1.0. Fitur time domain tersebut adalah standar deviasi, kurtosis, rms, maksimum absolut, faktor puncak, shape factor, dan impuls.

DATA SCIENTIST PROJECT

	Standar_Deviasi	Kurtosis	Root_Mean_Square_(RMS)	Maximum_Absolut	Faktor_Puncak	Faktor_Shape	Impuls	RUL
0	0.338628	0.000000	0.345119	0.165458	0.088994	0.000000	0.058878	50
1	0.183854	0.009311	0.183082	0.087879	0.088426	0.010971	0.059379	49
2	0.214218	0.010458	0.211309	0.146889	0.157828	0.004760	0.141482	48
3	0.144194	0.017901	0.141029	0.121683	0.179570	0.023968	0.163060	47
4	0.182692	0.019079	0.179568	0.133916	0.164164	0.021590	0.148871	46
5	0.198308	0.008740	0.200142	0.078570	0.031168	0.002405	0.026700	45
6	0.117734	0.008511	0.114561	0.046246	0.047261	0.004126	0.041411	44
7	0.000000	0.003957	0.000000	0.000000	0.065100	0.001517	0.057306	43
8	0.138324	0.029469	0.132873	0.138461	0.223911	0.038010	0.204657	42
9	0.126697	0.004515	0.125170	0.079591	0.107484	0.005323	0.095997	41
10	0.221175	0.050409	0.218093	0.191507	0.240114	0.080019	0.221620	40
11	0.236182	0.038195	0.231473	0.173261	0.189855	0.046352	0.174618	39
12	0.232406	0.066021	0.229503	0.243835	0.331854	0.069235	0.305922	38
13	0.324238	0.033604	0.334260	0.219350	0.176855	0.041793	0.162371	37
14	0.268939	0.062170	0.264024	0.304436	0.411831	0.075555	0.379304	36
15	0.274755	0.106783	0.270828	0.327756	0.449177	0.113451	0.417656	35
16	0.283700	0.137538	0.280038	0.373892	0.523140	0.163478	0.516378	34

Tabel di atas merupakan dataframe hasil normalisasi data dari fitur time domain yang terseleksi. Normalisasi data dilakukan agar data memiliki range nilai dari 0 sampai 1. Normalisasi data ini bertujuan agar model machine learning dapat menginterpretasikan data dengan lebih baik dari data dengan skala yang berbeda

```
print("data X train: ", X_train.shape)
print("data X test: ", X_test.shape)
print("data y train: ", y_train.shape)
print("data y test: ", y_test.shape)

data X train: (40, 7)
data X test: (10, 7)
data y train: (40,)
data y test: (10,)
```

Hasil dari train test split data dengan perbandingan 80:20.

Parameter	Nilai	
	Default	Hyperparameter Tuning
Learning rate	0.1	0.1
Max depth	3	2
N estimators	100	70

Konfigurasi parameter training dengan 2 parameter berbeda. Parameter pertama menggunakan parameter default, parameter kedua merupakan kombinasi nilai parameter hasil dari hyperparameter tuning dengan teknik grid search. Hyperparameter tuning menemukan kombinasi parameter dengan menurunkan max depth dan n estimators untuk mengurangi selisih antara RMSE train dan RMSE test serta untuk mengurangi risiko overfitting.

DATA SCIENTIST PROJECT

Model	RMSE Train (Hari)	RMSE Test (Hari)	Selisih
Model Default	0.1553	6.1040	5.9487
Model Tuning	1.6290	4.7624	3.1334

Tabel di atas menunjukkan bahwa terdapat perbedaan nilai RMSE train dan test pada masing-masing model yang menunjukkan adanya perbedaan performa model pada data train dan data test. RMSE test memiliki nilai yang lebih tinggi dibanding dengan RMSE train. Perbedaan ini normal terjadi karena adanya variasi alami dalam data. Ketika melatih model, model berusaha untuk mempelajari pola dan hubungan dalam data train untuk dapat memprediksi dengan akurat. Namun, karena variasi alami dalam data, tidak mungkin bagi model untuk mempelajari setiap keunikan yang ada dalam data train. Ketika model tersebut diuji pada data test dimana terdapat kemungkinan memiliki variasi yang berbeda dari data train, dapat memberikan performa yang berbeda antara RMSE train dan RMSE test. Perbedaan tersebut juga dapat menjadi indikasi adanya overfitting jika selisih antara RMSE train dan RMSE test besar. Namun, tidak ada rasio khusus atau angka pasti yang menentukan apakah model mengalami overfitting atau tidak.

Project ini dilakukan untuk mendapatkan nilai RMSE train dan RMSE test yang mendekati atau setidaknya tidak memiliki selisih yang signifikan serta memiliki nilai yang sekecil-kecilnya. Pada model default, RMSE test masih memperoleh angka 6.1040 hari dan memiliki selisih yang relatif tinggi dengan RMSE train. Dari model default ini dilakukan hyperparameter tuning agar RMSE test dapat menurun. Hyperparameter tuning dilakukan dengan metode grid search yang akan menemukan kombinasi parameter terbaik. Sesuai dengan teorinya, semakin tinggi nilai max depth dan n estimators maka risiko overfitting akan semakin tinggi. Oleh karena itu, hyperparameter tuning menemukan kombinasi parameter terbaik dengan menurunkan nilai max depth dan n estimators untuk mengurangi risiko terjadinya overfitting. Tuning ini berhasil menurunkan RMSE test dan menurunkan selisih antara RMSE train dan test. Oleh karena itu, model dengan tuning memiliki performa lebih baik dari model default.

DATA SCIENTIST PROJECT

Gambar di bawah merupakan hasil pembuatan sistem prediksi. Web dirancang untuk mengakses hasil model machine learning dengan lebih mudah. Perancangan web disusun sesuai kebutuhan. Fitur yang ada sebatas aplikasi prediksi RUL yang menjadi fokus dalam penelitian. Dalam pembuatan web, library pickle digunakan untuk menyimpan model machine learning. Penggunaan flask sebagai framework web atau kerangka kerja back-end, HTML digunakan sebagai bahasa markup untuk membuat tampilan halaman antar muka, dan CSS untuk mengatur gaya dan tata letak elemen-elemen pada halaman antar muka.

Melalui sistem ini, pengguna dapat memasukkan file rekaman vibrasi dari direktori lokal yang berformat matlab (.mat) melalui button upload section. Data yang dimasukkan kemudian dikirim ke server flask yang mengelola pemrosesan dan prediksi RUL menggunakan model yang telah ditraining sebelumnya. Hasil prediksi RUL kemudian ditampilkan kembali ke pengguna melalui halaman web yang menggunakan HTML untuk mengatur tampilan dan struktur data serta CSS untuk memberikan gaya dan tata letak tampilan.

RUL BEARING's PREDICTION

Upload File:

No file chosen

Prediction: 35 Day left

Time Domain Information:

Standar_Deviasi: 0.188747

Kurtosis: 0.044121

Root_Mean_Square_(RMS): 0.18474

Maximum_Absolut: 0.213575

Faktor_Puncak: 0.321422

Faktor_Shape: 0.048895

Impulse: 0.294241

© 2023 [Nurkhanifah] - Tugas Akhir

CONCLUSION

Berdasarkan project yang dilakukan beserta pengujian yang telah dilakukan, dapat disimpulkan bahwa model prediksi RUL yang dibangun menggunakan algoritma gradient boosting memiliki performa yang lebih baik daripada penelitian sebelumnya, yaitu menggunakan algoritma SVR. Hal ini dibuktikan oleh nilai RMSE terbaik dengan algoritma gradient boosting sebesar 4.7624, sedangkan RMSE terbaik yang dihasilkan dengan algoritma SVR dengan perlakuan yang sama terhadap data dan fitur yang sama adalah sebesar 8.5927. Perolehan angka tersebut dapat memberikan kesimpulan bahwa algoritma gradient boosting memiliki kinerja yang lebih baik dari pada algoritma SVR