Model Prediction for Machine Failure

The Game Changers Team

Nouval Habibie

Mentor

Anggun Dwi Cahyani

Facilitator

I Gede Purna Adi Putra

Universitas Udayana

Ivanna Ruth Theophilia Putri

Universitas Brawijaya

Kharir Nur Asniyah

Universitas Diponegoro

Lukman

Universitas Pendidikan Indonesia

Mesy Diah Ariani

Universitas Gajah Mada

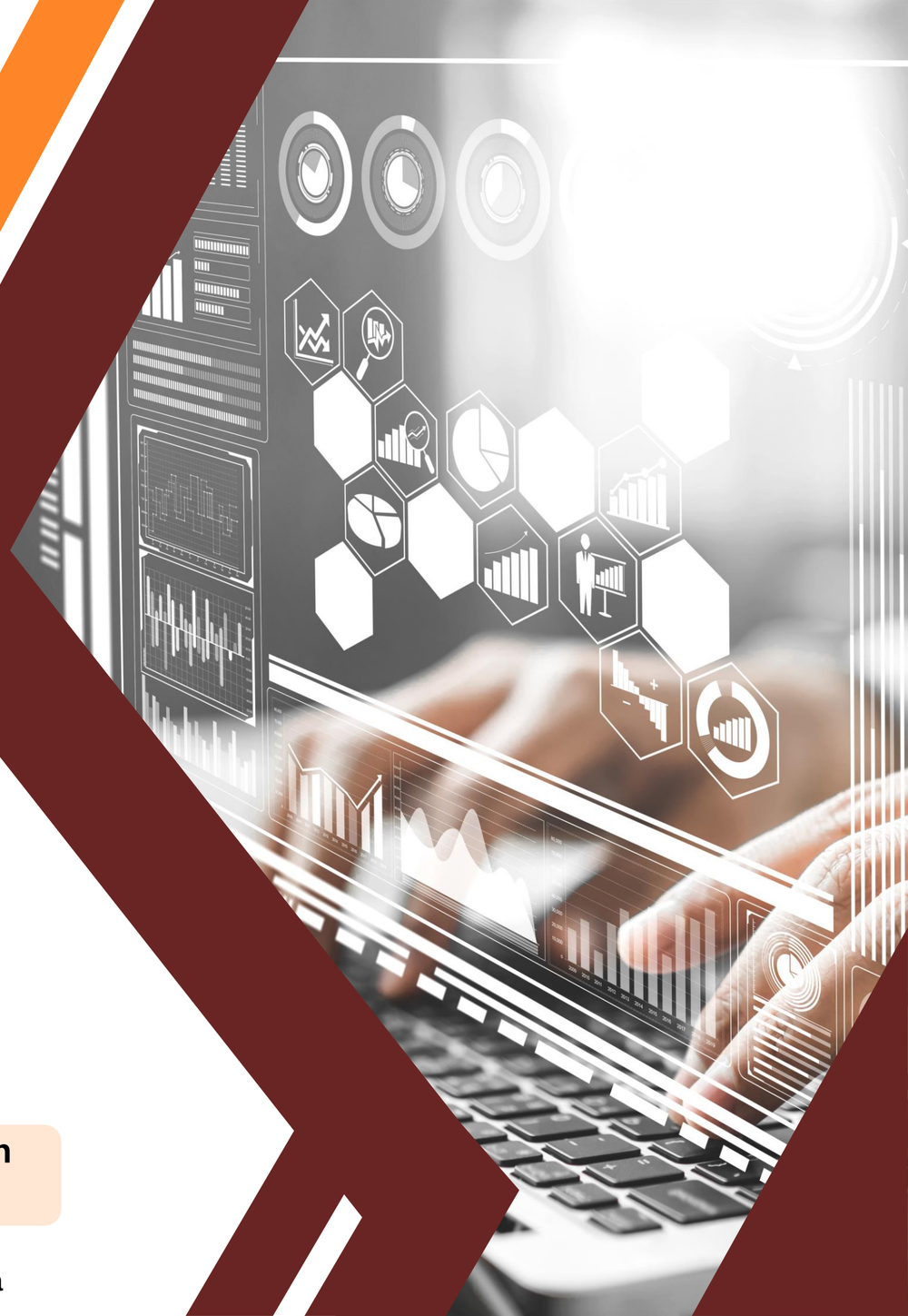
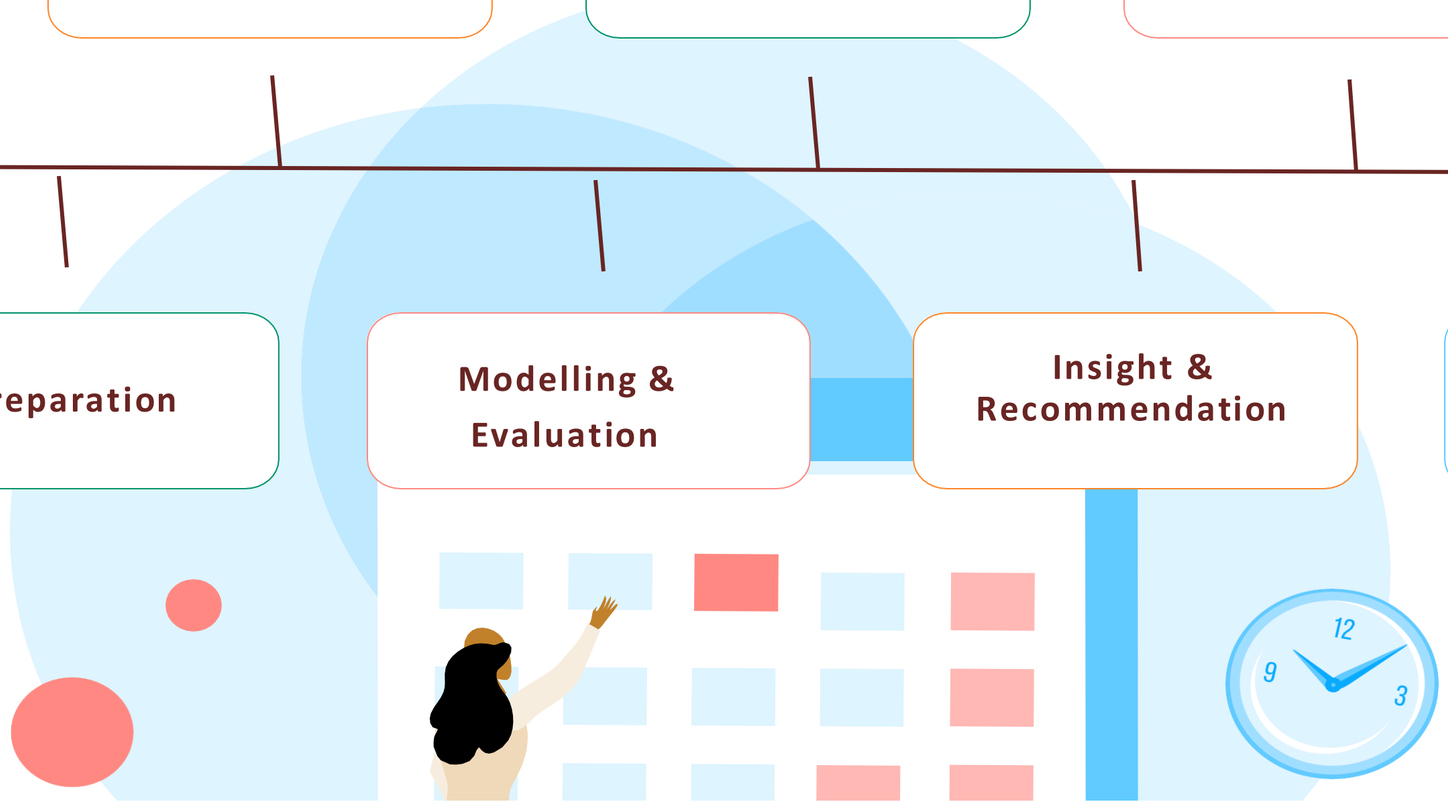


Table of Content



Problem

Sebuah perusahaan sparepart bereputasi baik

Ingin melakukan ekspansi

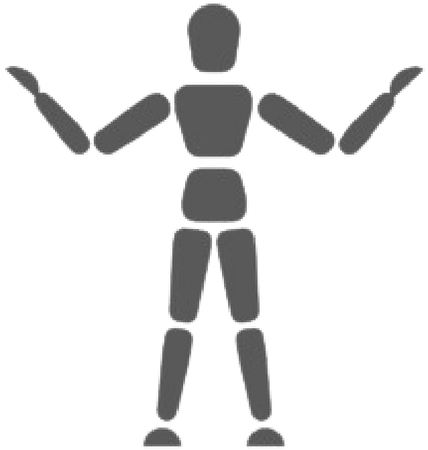
Muncul masalah -> Sering terjadi kerusakan mesin

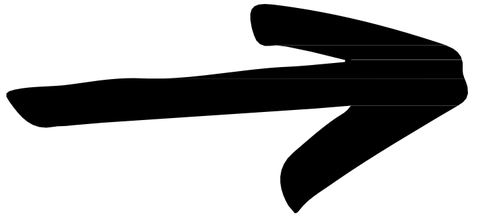
Muncul biaya tambahan

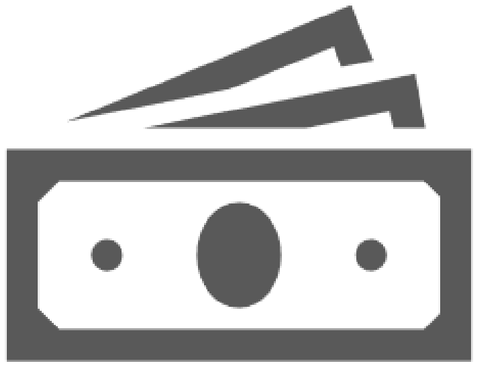
Goals

- Mengetahui penyebab umum kegagalan mesin.
- Meninjau biaya terkait kegagalan mesin.
- Mengidentifikasi peluang penghematan biaya.

Solution Framework



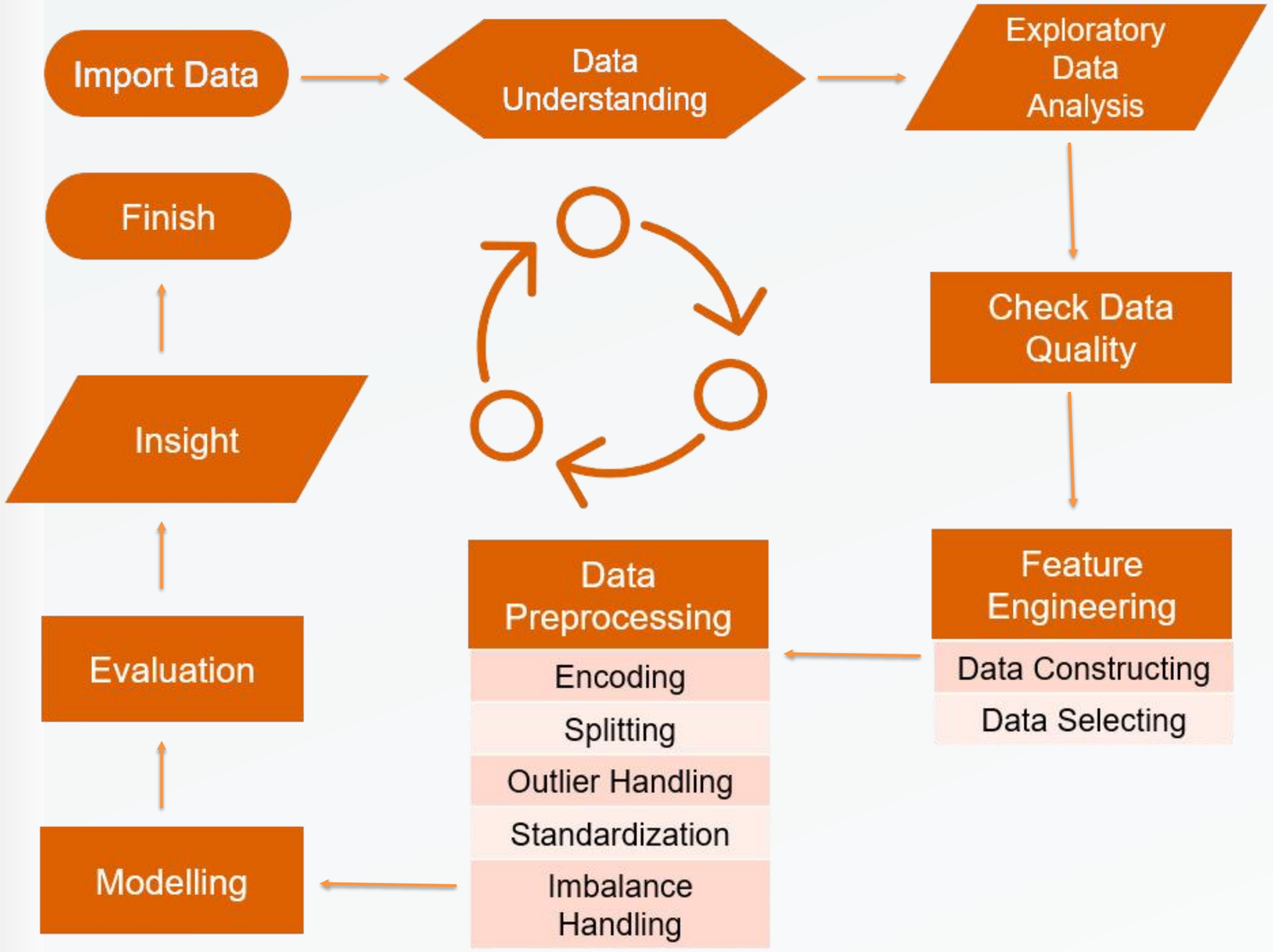




Pemodelan

Dapat menghemat pengeluaran

Analysis Stage



Data Understanding

Variabel	Tipe Data	Keterangan
UDI	Nominal	Pengenal unik yang berkisar dari 1 hingga 10.000
Product ID	Nominal	Terdiri dari huruf L, M, atau H serta nomor seri yang spesifik untuk setiap varian.
Type	Nominal	Tipe produk terdiri dari huruf L, M, atau H
Air temperature	Kontinu	Dihasilkan menggunakan proses random walk yang kemudian dinormalisasi dengan standar deviasi 2 K di sekitar 300 K.
Process temperature	Kontinu	Dihasilkan menggunakan proses random walk yang dinormalisasi dengan standar deviasi 1 K, ditambah dengan suhu udara ditambah 10 K.
Rotational speed [rpm]	Kontinu	Dihitung dari daya sebesar 2860 W, dengan penambahan noise yang terdistribusi secara normal
Torque [Nm]	Kontinu	Nilai torsi terdistribusi secara normal sekitar 40 Nm dengan SD = 10 Nm dan tidak ada nilai negatif.
Tool wear [min]	Kontinu	Variasi kualitas H/M/L menambahkan waktu keausan alat sebesar 5/3/2 menit pada alat yang digunakan.
Machine failure	Nominal	Kegagalan mesin pada titik tertentu karena salah satu dari lima mode kegagalan independen.
TWF	Nominal	Tool wear failure
HDF	Nominal	Heat dissipation failure
PWF	Nominal	Power failure
OSF	Nominal	Overstrain failure
RNF	Nominal	Random failures

Column : 14
Row : 10.000
Source : https://archive.ics.uci.edu/ml/datasets/AI4I+2020+Predictive+Maintenance+Dataset#

Data Preparation

1 DATA QUALITY

Tidak ada data missing value dan data duplikat.

2 DATA CONSTRUCTING

- Kategori ulang variabel *machine failure*
- Membuat fitur baru dari perkalian Torque & Rotational Speed yaitu, "**Power**".

3 DATA SELECTING

Drop variabel UDI, Product ID, PWF, HDF, OSF, TWF, RNF

4 DATA ENCODING

Encoding variabel "**Type**"

5 SPLITTING DATASET

TEST SIZE	TRAIN SIZE
0.25	0.75

6 OUTLIER HANDLING

- Outlier pada variabel Rotational Speed, Torque, & Power.
- Ditangani menggunakan Box-Cox Transformation.

7 STANDARDIZATION

Robust Scaler Standardization

8 IMBALANCE DATA HANDLING

SMOTENN Hybrid Technique

No	Algoritma	Kelas	Precision	Recall	F1-score	Akurasi	AUC
1	Logistic Regression	0	0,99	0,87	0,92	0,86	0,87
	Hyperparameter tuning	1	0,14	0,71	0,24		
2	Support Vector Machine	0	1	0,93	0,96	0,93	0,95
	Baseline	1	0,29	0,88	0,43		
3	Decision Tree	0	0,99	0,92	0,96	0,92	0,89
	Hyperparameter tuning	1	0,25	0,82	0,38		
4	Random Forest	0	1	0,92	0,96	0,92	0,95
	Hyperparameter tuning	1	0,27	0,88	0,41		
5	Gradient Boosting	0	0,99	0,97	0,98	0,96	0,95
	Hyperparameter tuning	1	0,45	0,79	0,57		
6	XGBoost	0	0,99	0,96	0,98	0,96	0,94
	Hyperparameter tuning	1	0,41	0,84	0,55		

Modelling & Evaluation

Pemodelaan dengan **algoritma XGBoost** memiliki metrics evaluation yang paling stabil.

Insight


Animal feed
Food Business
Farm business
total
1,325,734.00
321,005.00
100.00
1,502,231.00
386,457.00
524,987.00
590,787.00
39.33
34.95
25.73
100.00
622,400.00
573,258.00
421,897.00
1,617,555.00
38.48
35.44
26.08
100.00
Food Business, 36.85
Farm business, 24.21
Animal feed, 38.94
total , 100.00
Food Business, 57.70
Animal feed, 9.60
Food Business 37%

Cost & Benefit based on Confusion Matrix

	Cost	Keterangan
False Positive	*x*	Unnecessary maintenance
False Negative	20*x*	Machine runs to failure
True Positive	5*x*	Timely and appropriate maintenance
True Negative	0	Has no cost because no action taken

Diasumsikan bahwa, jika mesin diprediksi gagal padahal sebenarnya tidak gagal (False Positif) akan mengeluarkan biaya sebesar x dollar.

Cost	FP	FN	TP	TN	Total Cost (Dollar)	Jumlah mesin
Baseline Cost	1224	38	38	1200	2174*x*	2500

Biaya Historis perbaikan dan maintenance mesin dapat diperoleh dengan simulasi distribusi uniform.

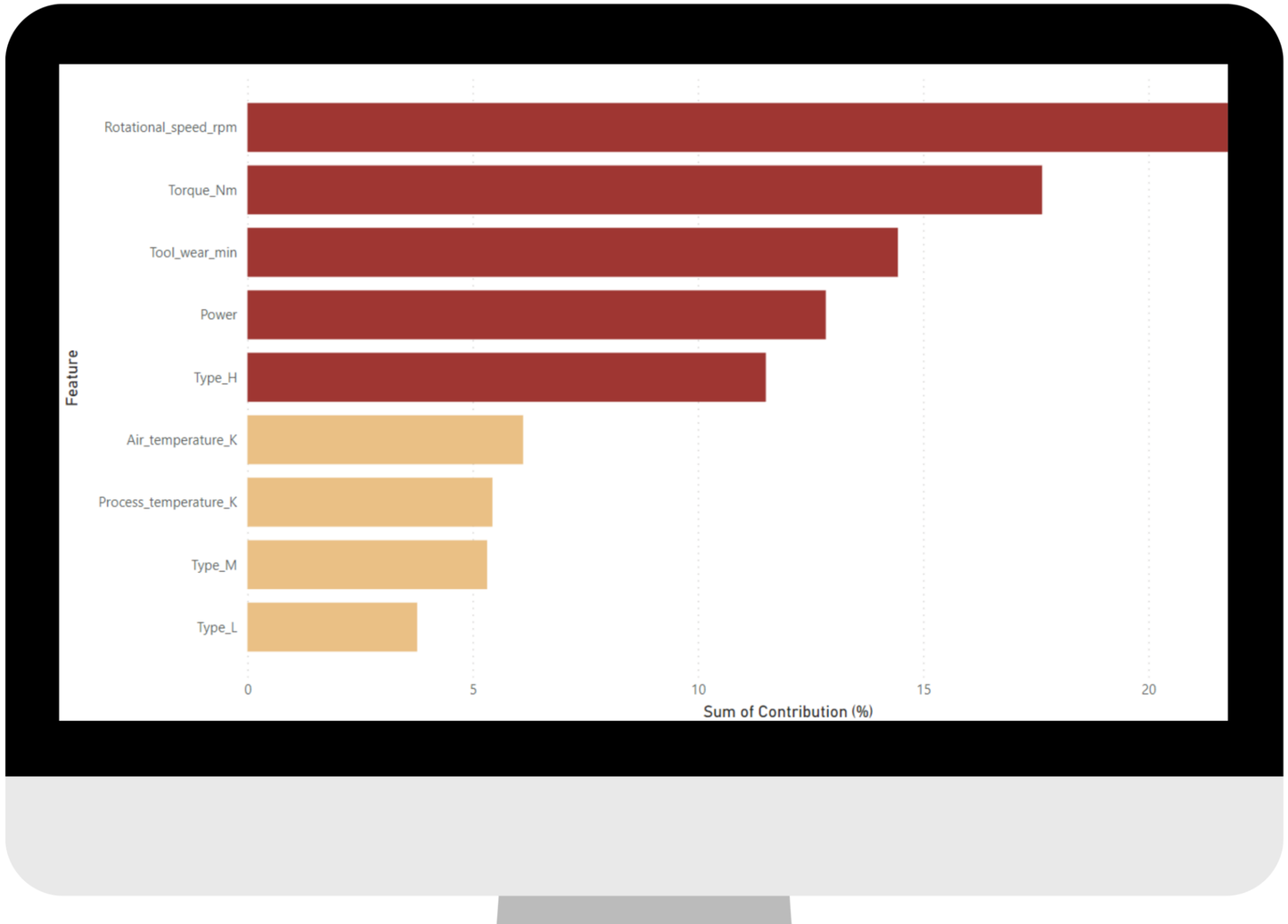
Cost & Benefit based on Confusion Matrix

No	Algoritma	FP	FN	TP	TN	Total Cost (Dollar)	Jumlah mesin
1	Logistic Regression	320	22	54	2104	1030*x*	2500
2	Support Vector Machine	168	9	67	2256	683*x*	2500
3	Decision Tree	187	14	62	2237	777x	2500
4	Random Forest	182	9	67	2242	697*x*	2500
5	Gradient Boosting	73	16	60	2351	693*x*	2500
6	XGBoost	93	12	64	2331	653*x*	2500

Cost Efficiency		
$\frac{\text{Total cost baseline}}{\text{Total cost XGBoost}}$	=	3,329

Dengan model terbaik **XGBoost** dapat menghemat biaya sebesar **3,33 kali lipat** dari biaya sebelumnya.

Feature Importance



Top 5 Feature Importance:

- Rotational Speed
- Torque
- Tool Wear
- Power
- Type H

Recommendation

Mengimplementasikan jadwal perawatan rutin.

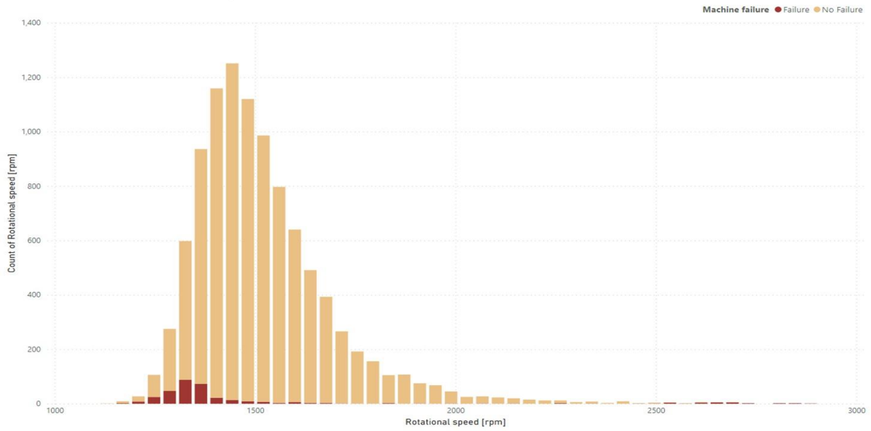
Memantau dan mendeteksi tanda awal kegagalan mesin.

Fokus pada peningkatan keandalan feature importance.

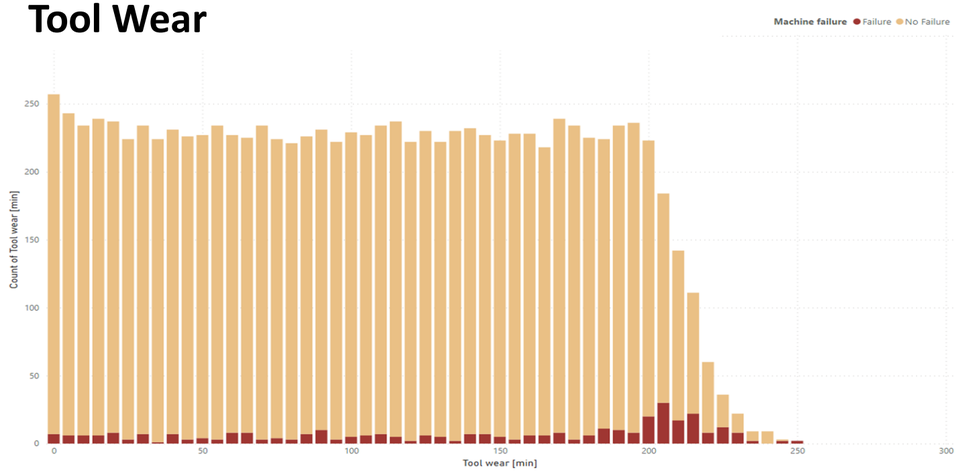
Bekerja sama dengan pemasok suku cadang berkualitas tinggi.

Lampiran EDA

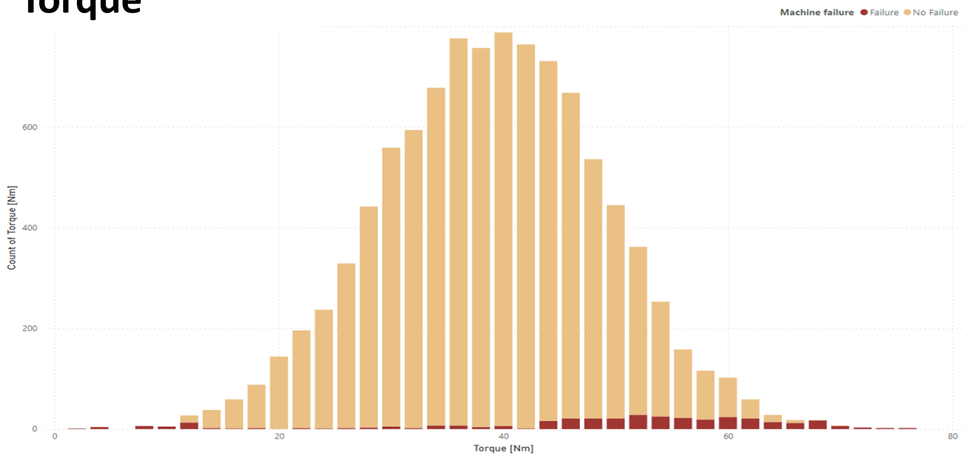
Rotational Speed



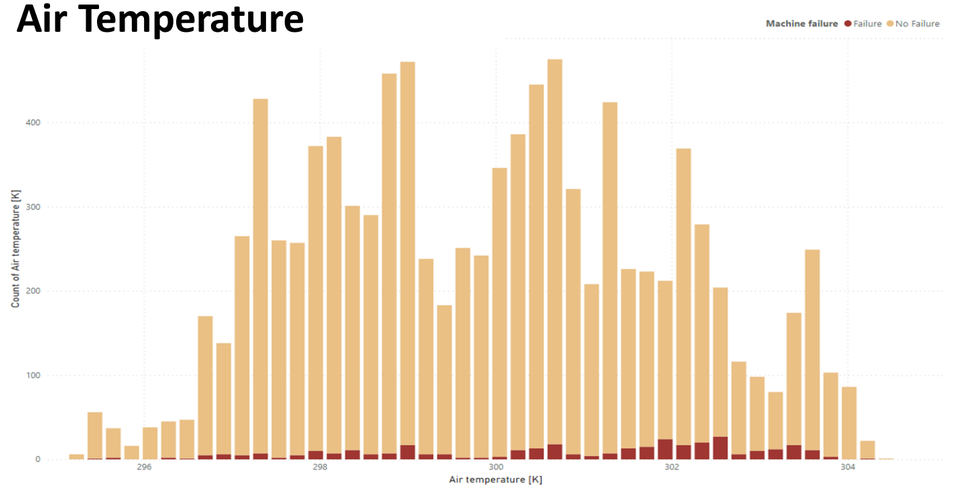
Tool Wear



Torque

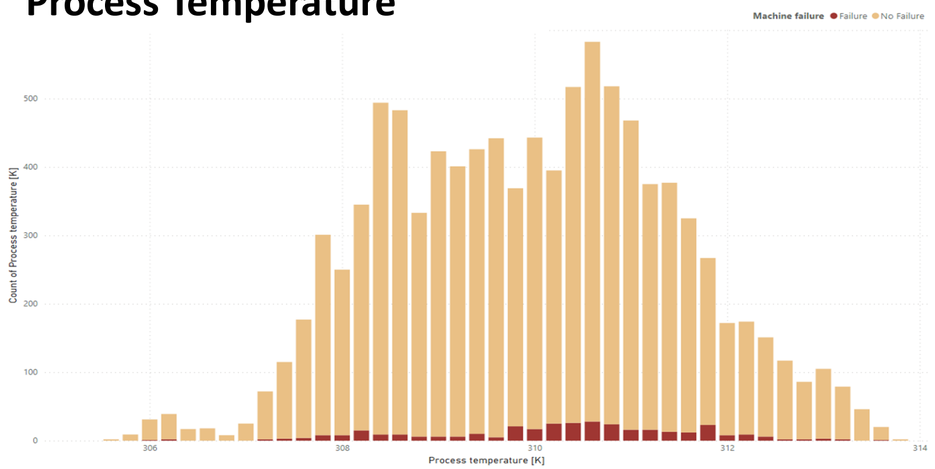


Air Temperature

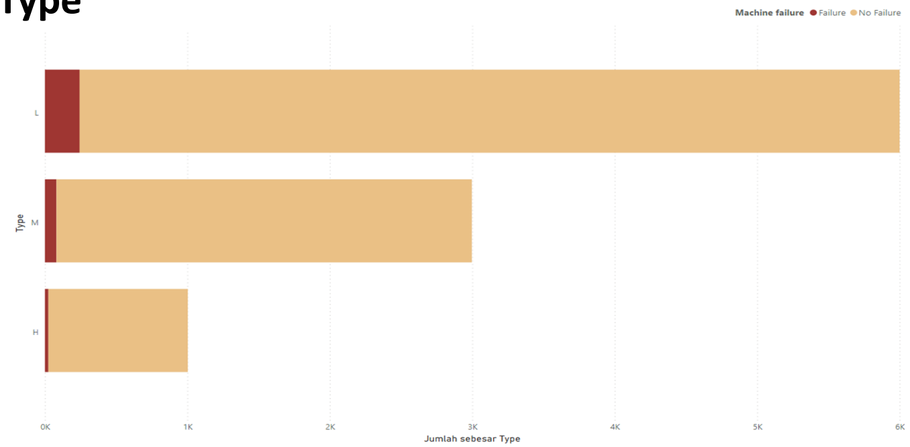


Lampiran EDA

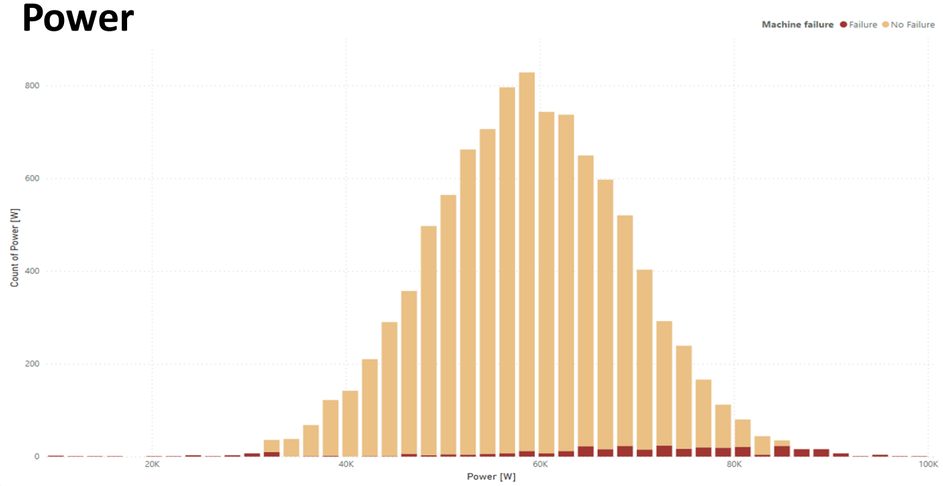
Process Temperature



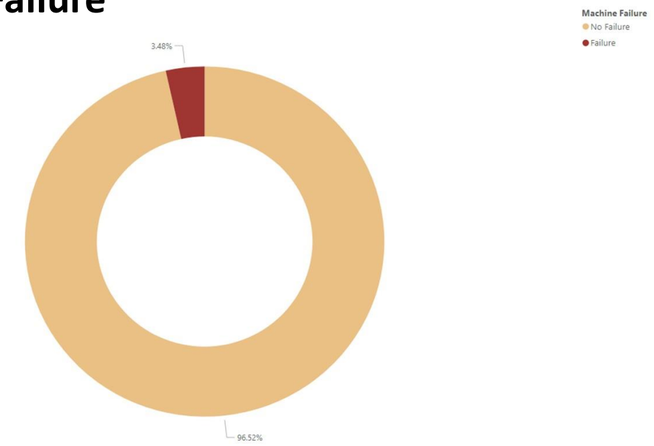
Type



Power



Machine Failure



Lampiran Dashboard

Overview

Analysis

Select

Machine Failure

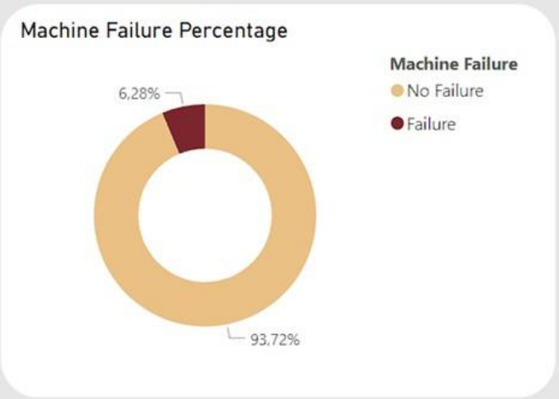
Semua

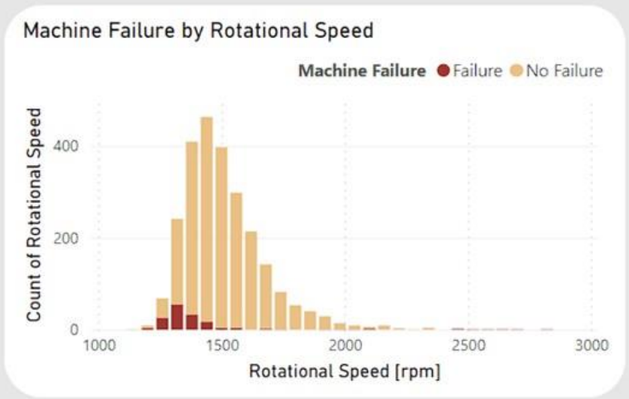
Type

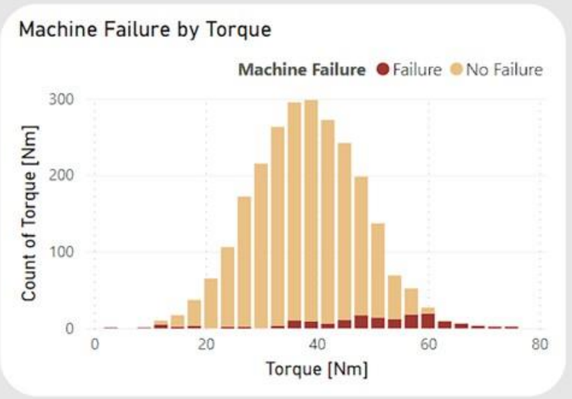
Semua

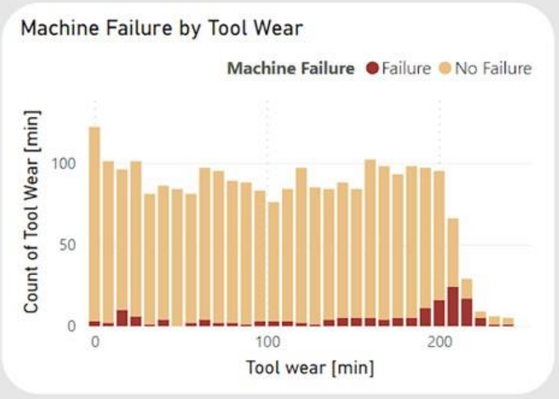
Machine Failure Overview

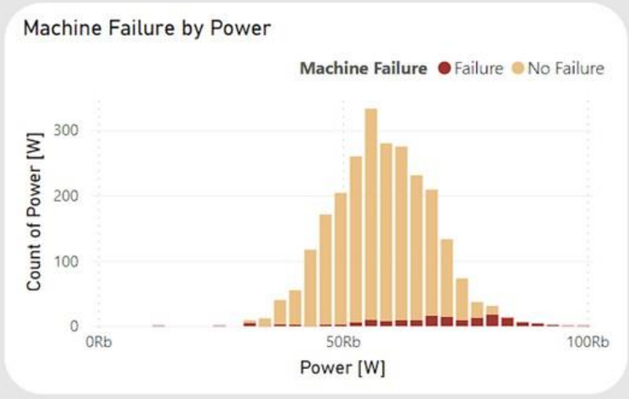
Total Machine	Avg Rot. Speed	Avg Torque	Avg Tool Wear	Avg Power	Failure Probability
2500	1,54Rb	39,73	108,03	59,73Rb	0,00 – 1,00

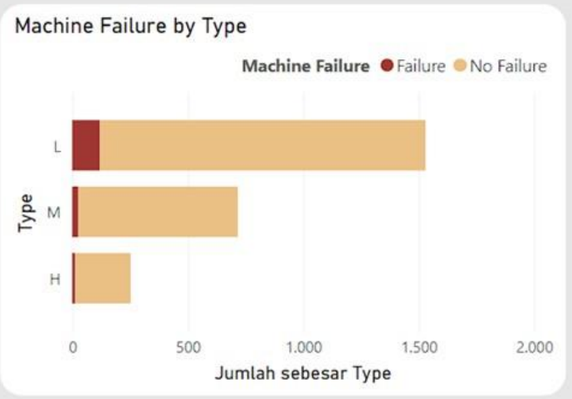












Lampiran Dashboard

Overview

Analysis

XGboost Model Analysis

Summary Test Result

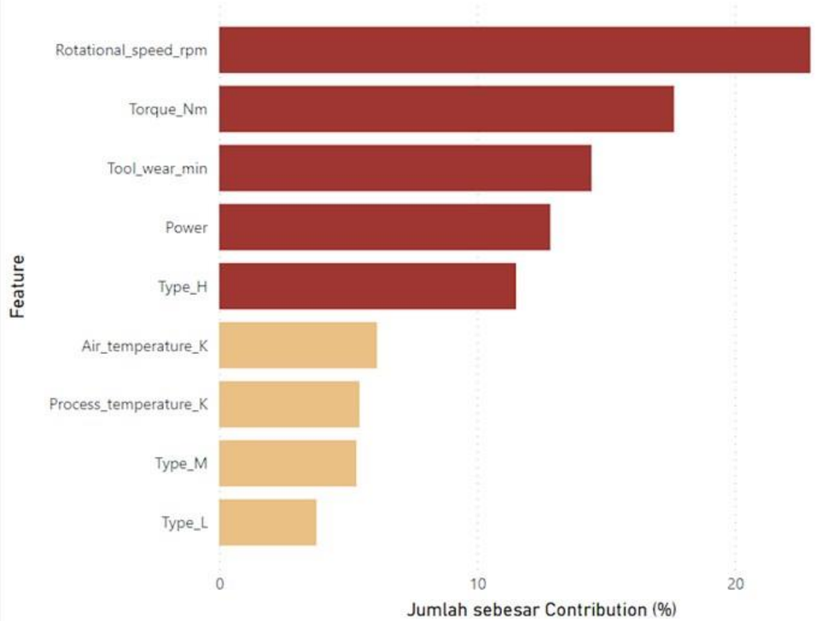
Area Under Curve

Recall

Confusion Matrix

Actual/Prediction	Failure	No Failure	Total
Failure	64	12	**76**
No Failure	93	2331	**2424**
Total	**157**	**2343**	**2500**

Feature Importance



Cost Benefit Analysis

Previous Cost	XGboost Cost	Cost Efficiency
2174	653	3,33
In the repair and maintenance of machine, the company spent 2174 units x dollars	In the repair and maintenance of machine, the company spent 653 units x dollars if the company uses this model	This model will save cost 3.33 times than the previous cost

Insight

- With a 22.93% contribution, rotational speed is the main cause of engine failure.
- Torque contributed to affecting engine failure by 17.64%
- Tool wear contributed to affecting engine failure by 14.44%.
- Power contributed to affecting engine failure by 12.84%.
- Engines with Type H contributed to engine failure by 11.51%.
- Company can save cost 3.33 times than the previous cost.

Recommendation

- The company needs to maintain rotational speed, torque, tool wear, and power on the machine in ideal conditions to prevent engine failure.
- Machines with type H have a higher risk of failure than other types of machines, so businesses should periodically check these devices.
- It is more cost-effective for the business to do routine maintenance on the machine to avoid machine failure than it is to fix the machine that has already failed. Testing of rotational speed, torque, tool wear,

Terima Kasih

https://bit.ly/Coding_The_Game_Changers

https://bit.ly/Dashboard_The_Game_Changers