Machine Learning Based Pipeline Integrity Assessment for Predictive Maintenance

Bagus Cipta Pratama

2026

Data yang ada pada presentasi ini merupakan data modifikasi yang dimiliki PT Kilang Pertamina Internasional dan presentasi ini dibuat untuk portfolio pribadi

Latar Belakang

Pipeline merupakan aset kritis dalam industri minyak dan gas yang rentan mengalami degradasi akibat korosi, tekanan, dan umur operasi. Pendekatan pemeliharaan konvensional belum sepenuhnya mempertimbangkan kondisi aktual pipa.

Oleh karena itu, pemanfaatan data inspeksi dan machine learning digunakan untuk membangun model predictive maintenance yang mampu mengklasifikasikan kondisi pipeline secara akurat dan mendukung pengambilan keputusan pemeliharaan berbasis risiko.

Tujuan Bisnis

1. Mengklasifikasikan kondisi pipeline berdasarkan data inspeksi,
2. Mengidentifikasi pipeline dengan risiko tinggi secara lebih dini,
3. Mendukung pengambilan keputusan pemeliharaan yang lebih efektif dan efisien
4. Mengurangi potensi downtime, kegagalan, dan biaya perbaikan mendadak

Dataset Overview

Dataset yang digunakan berasal dari data inspeksi pipeline yang mencakup informasi kondisi fisik dan operasional pipa. Dataset ini memiliki **1000 baris dan 11 kolom**,

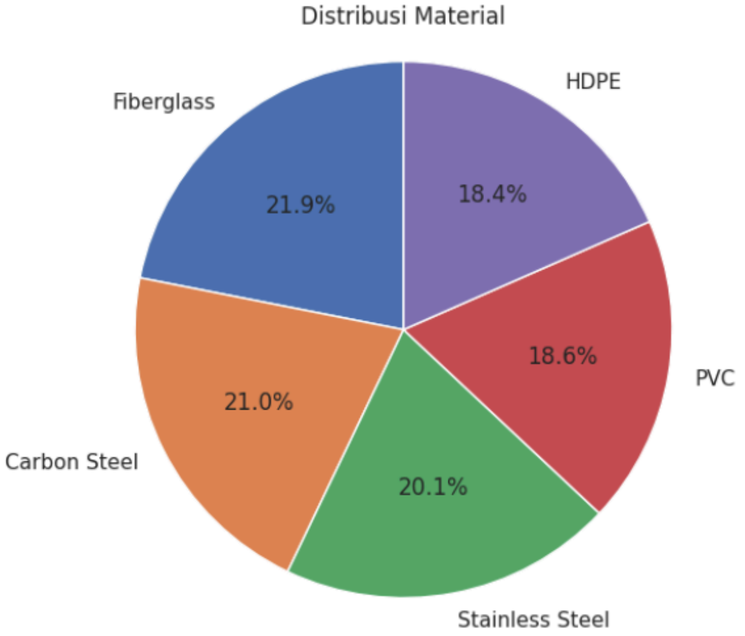
dengan fitur utama seperti **ketebalan pipa, kehilangan ketebalan, tekanan, temperatur, umur operasi, serta karakteristik material**.

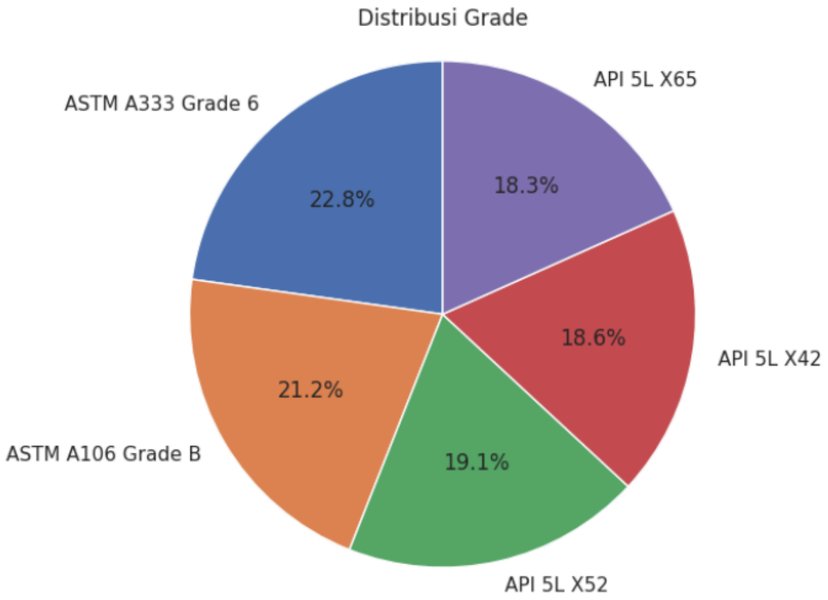
Variabel target pada dataset ini adalah **Condition**, yang merepresentasikan kondisi pipeline dan digunakan sebagai dasar klasifikasi dalam model predictive maintenance.

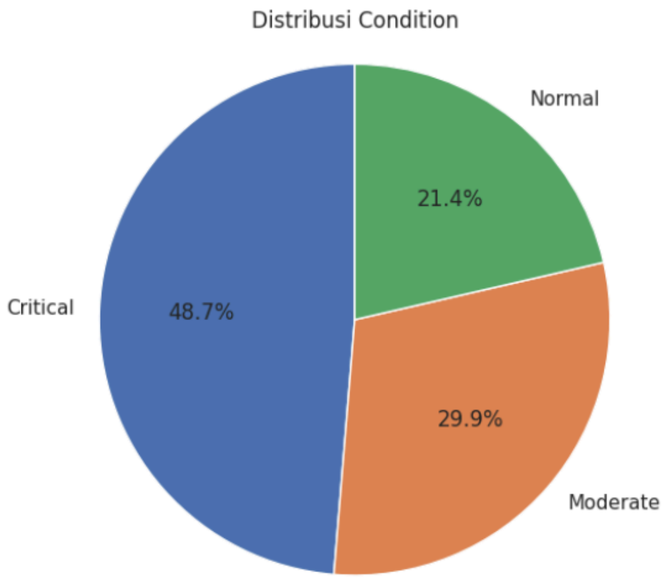
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 11 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Pipe_Size_mm               1000 non-null   int64
 1   Thickness_mm               1000 non-null   float64
 2   Material                   1000 non-null   object
 3   Grade                      1000 non-null   object
 4   Max_Pressure_psi           1000 non-null   int64
 5   Temperature_C              1000 non-null   float64
 6   Corrosion_Impact_Percent   1000 non-null   float64
 7   Thickness_Loss_mm          1000 non-null   float64
 8   Material_Loss_Percent      1000 non-null   float64
 9   Time_Years                 1000 non-null   int64
 10  Condition                  1000 non-null   object
dtypes: float64(5), int64(3), object(3)
memory usage: 86.1+ KB
```

Data Understanding

setelah dilihat dan diobservasi , data memiliki nol missing value dan nol duplikat yang mencerminkan record yang cukup bersih dan konsisten . berikut adalah distribusi kolom kategorikal pada dataset .





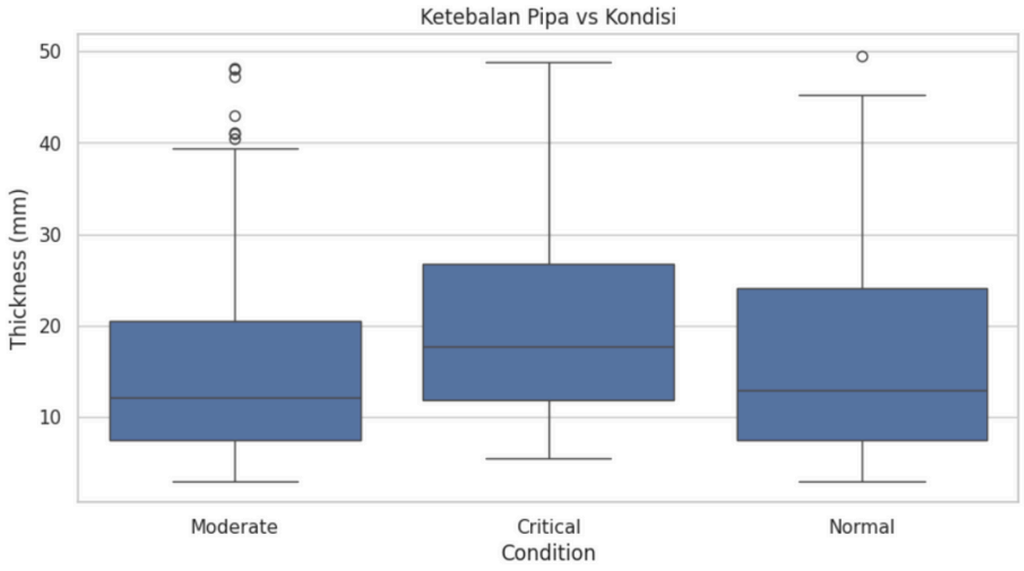


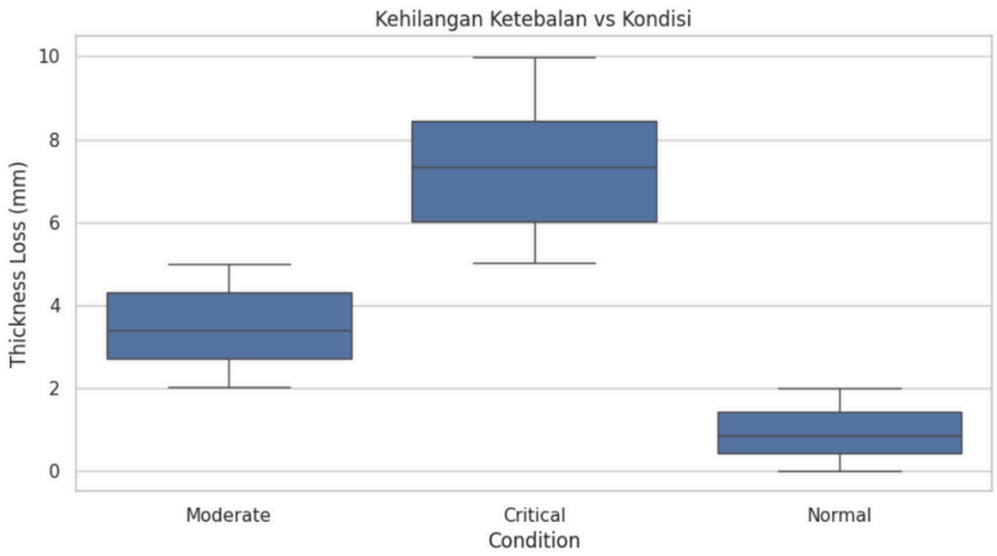
Data Cleaning & Validation

- Ketebalan pipa harus bernilai positif
- Kehilangan ketebalan tidak boleh negatif
- Kehilangan ketebalan tidak melebihi ketebalan awal
- Umur pipa tidak boleh negatif

setelah melakukan cleaning dan tahap validasi , data bersih yang siap diproses untuk tahap akhir berjumlah 881 baris data

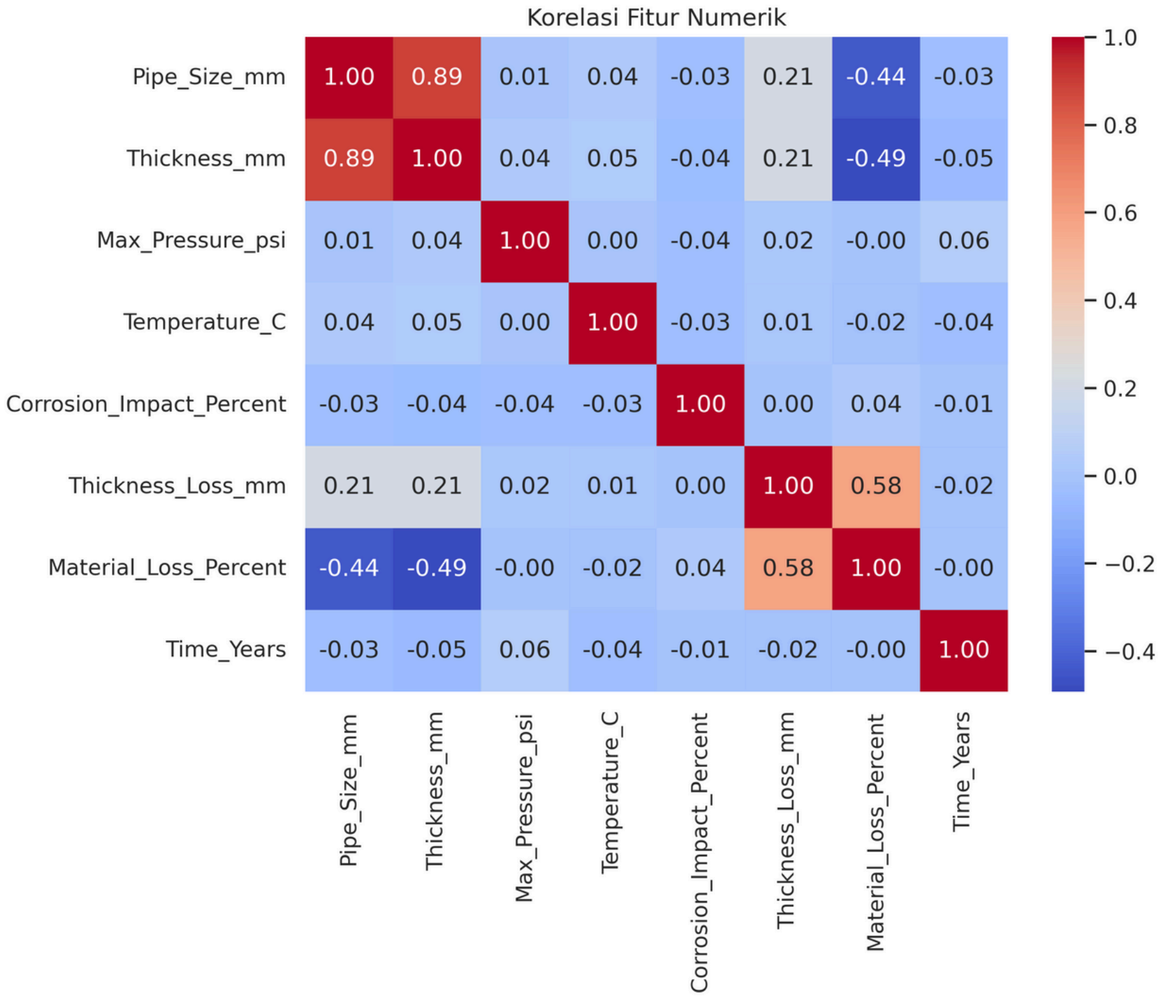
Exploratory Data Analysis 1

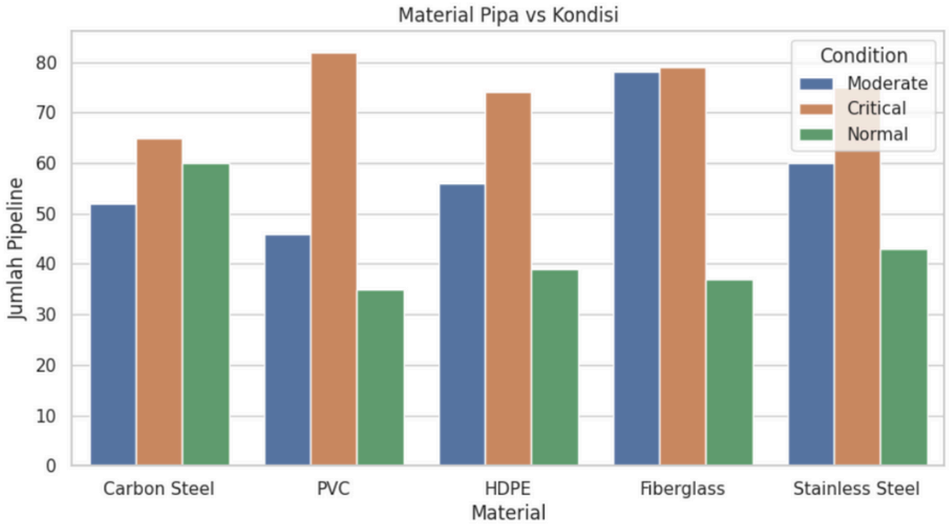


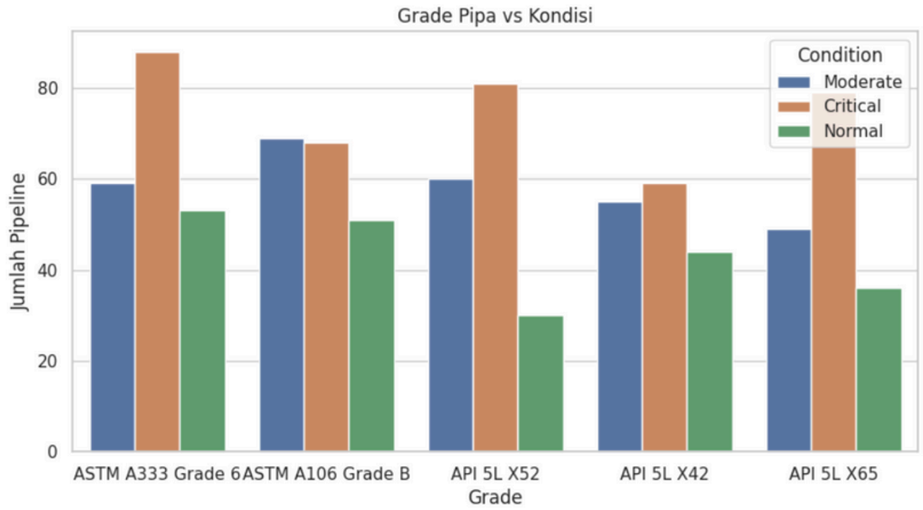


Kehilangan ketebalan pipa **meningkat secara signifikan** pada pipeline dengan kondisi Critical, menjadikannya indikator utama degradasi, sementara ketebalan awal pipa berperan sebagai faktor mitigasi di mana pipa yang lebih tebal **cenderung memiliki kondisi yang lebih baik.**

Exploratory Data Analysis 2







Exploratory Data Analysis 2

- Material dan grade pipa **memengaruhi tingkat risiko**, namun bukan faktor tunggal penentu kondisi.
- Kehilangan material dan kehilangan ketebalan memiliki **hubungan yang kuat** dan menjadi indikator degradasi utama.
- Fitur operasional dan dimensi fisik memiliki **korelasi lemah** secara individual, menegaskan perlunya pendekatan multivariat.

Feature Engineering & Preprocessing

- Dibuat fitur rasio kehilangan ketebalan untuk merepresentasikan tingkat degradasi pipa secara relatif terhadap ketebalan awal.

```
# Rasio kehilangan ketebalan terhadap ketebalan awal
X_fe["Thickness_Loss_Ratio"] = (
    X_fe["Thickness_Loss_mm"] / X_fe["Thickness_mm"]
)

# Pastikan tidak ada inf atau NaN akibat pembagian
X_fe["Thickness_Loss_Ratio"] = X_fe["Thickness_Loss_Ratio"].replace(
    [np.inf, -np.inf], np.nan
)
```

Feature Engineering & Preprocessing

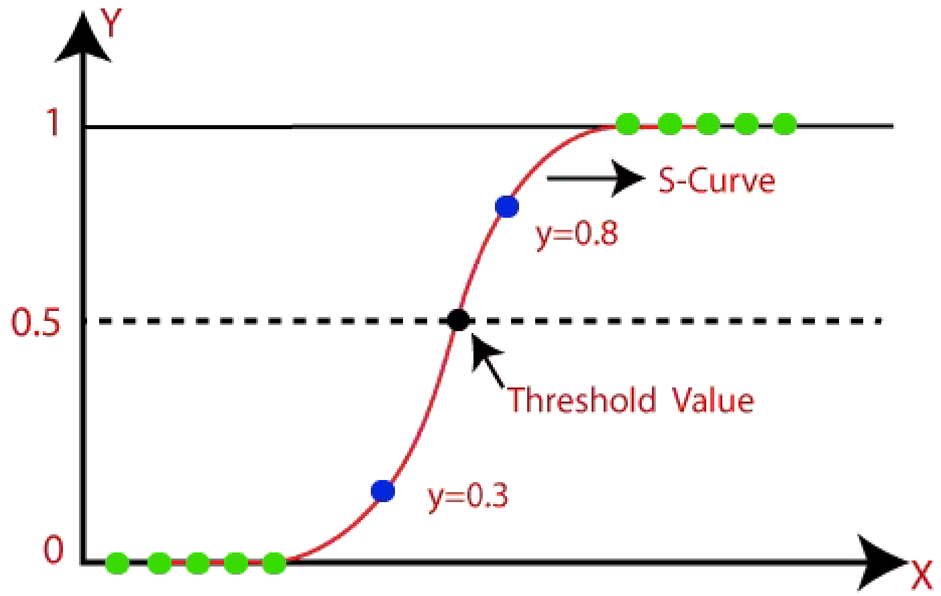
- Fitur kategorikal seperti material dan grade pipa diubah menjadi format numerik menggunakan encoding agar dapat diproses oleh model machine learning.
- Dataset dibagi menjadi data latih dan data uji, kemudian fitur numerik dilakukan scaling untuk memastikan proses training berjalan stabil dan tanpa data leakage.

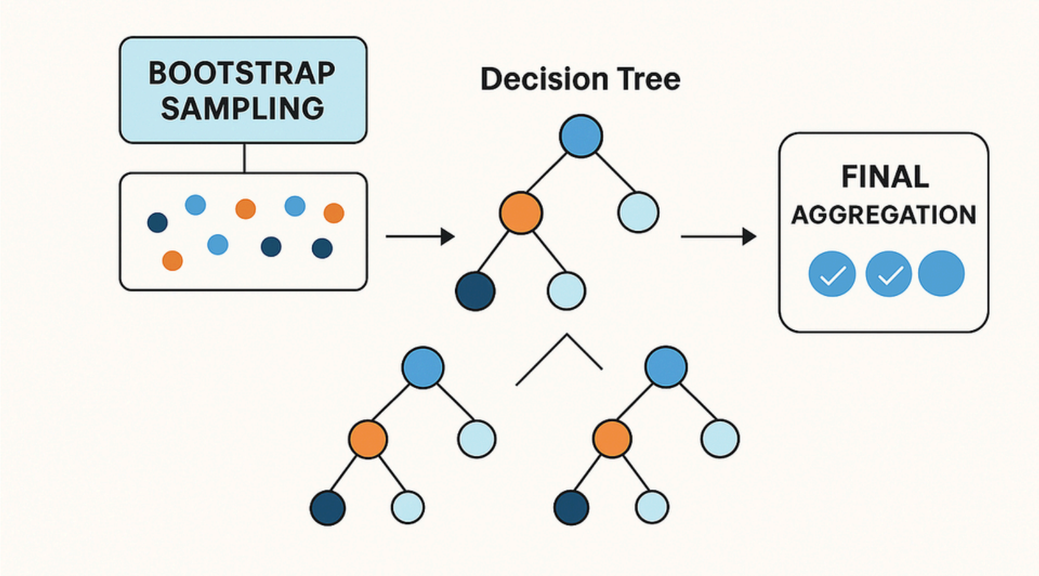
```
X_encoded = pd.get_dummies(
    X_fe,
    columns=categorical_features,
    drop_first=False
)
```

```
X_train, X_test, y_train, y_test = train_test_split(
    X_encoded,
    y_encoded,
    test_size=0.2,
    random_state=RANDOM_STATE,
    stratify=y_encoded
)
```

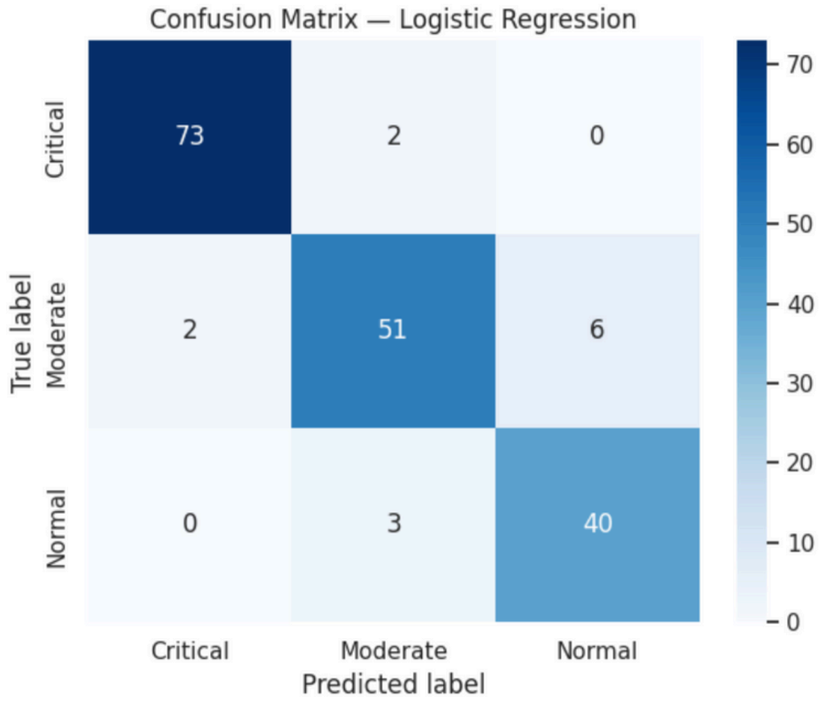
Modelling Approach

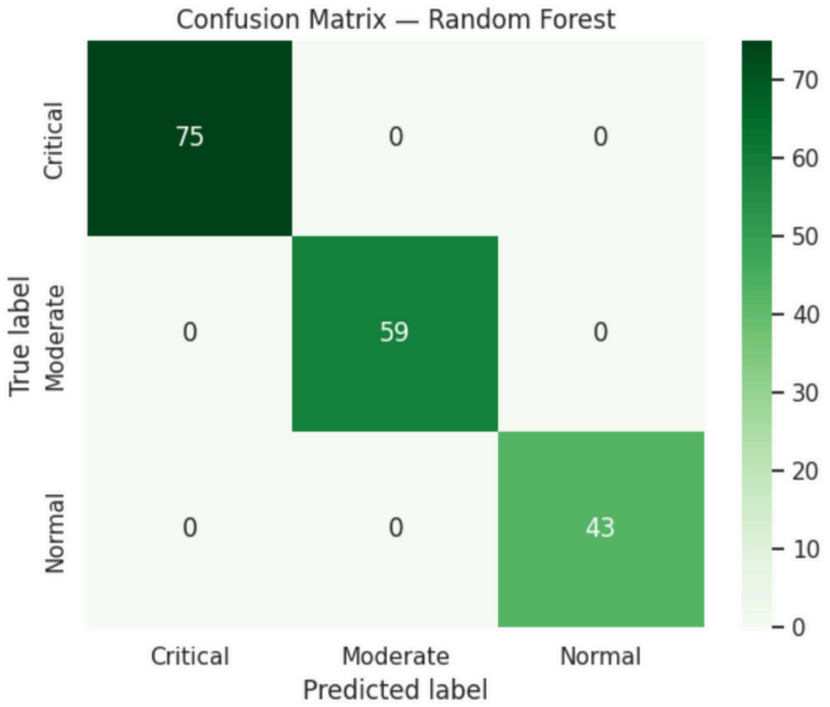
Pendekatan pemodelan dilakukan dengan membandingkan model baseline **Logistic Regression** dan model utama **Random Forest** untuk menangkap hubungan non-linear dalam data serta memilih model dengan performa terbaik dalam mengklasifikasikan kondisi pipeline.





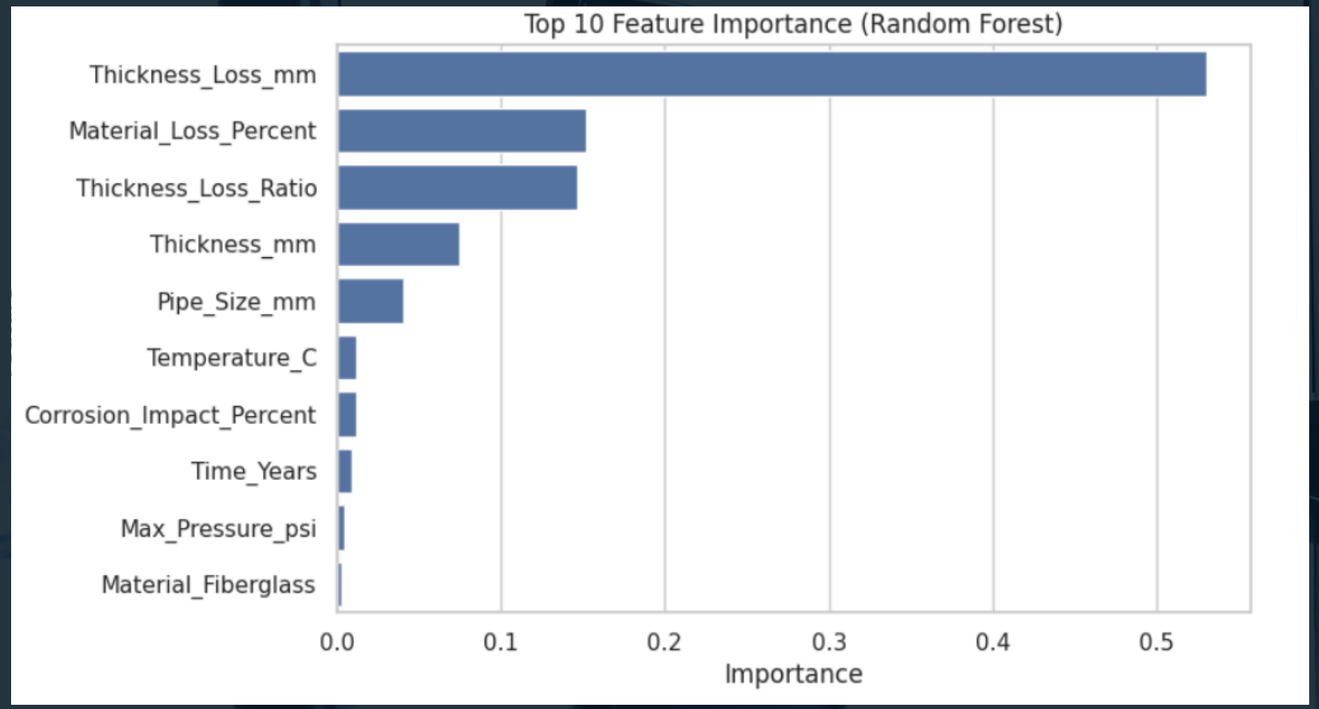
Modelling Evaluation





	Model	Accuracy	Macro Precision	Macro Recall	Macro F1
0	Logistic Regression	0.9266	0.9179	0.9227	0.9197
1	Random Forest	1.0000	1.0000	1.0000	1.0000

Model Interpretability

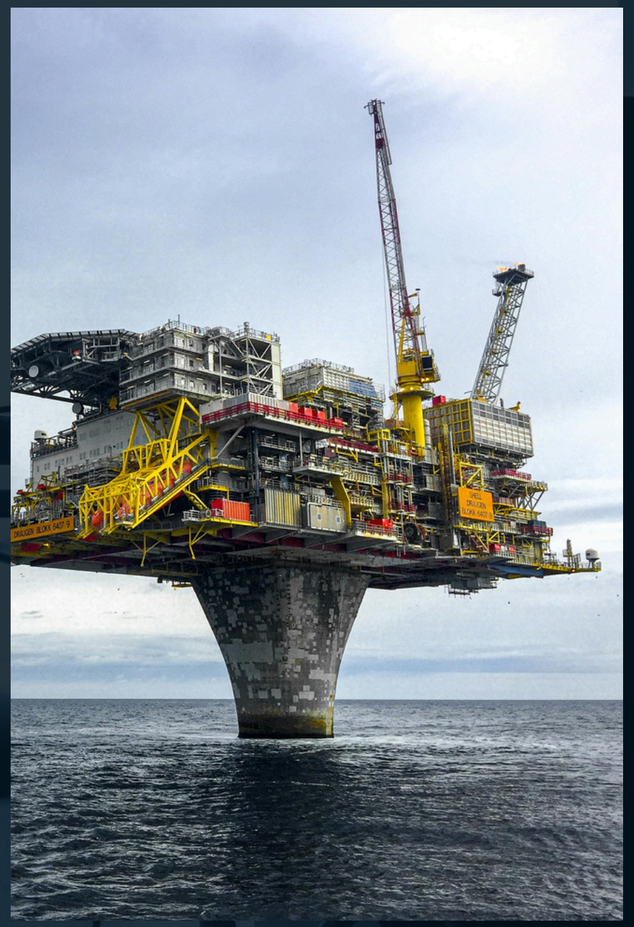


	feature	importance
5	Thickness_Loss_mm	0.5303
6	Material_Loss_Percent	0.1523
8	Thickness_Loss_Ratio	0.1463
1	Thickness_mm	0.0753
0	Pipe_Size_mm	0.0407
3	Temperature_C	0.0124
4	Corrosion_Impact_Percent	0.0123
7	Time_Years	0.0095
2	Max_Pressure_psi	0.0052
10	Material_Fiberglass	0.0026

Hasil feature importance menunjukkan bahwa kehilangan **ketebalan pipa merupakan faktor paling dominan** dalam menentukan kondisi pipeline, **diikuti oleh kehilangan material dan rasio kehilangan ketebalan**, sementara faktor operasional seperti tekanan, temperatur, dan umur memiliki kontribusi yang relatif lebih kecil.

Conclusion & Recommendation

- **Kehilangan ketebalan pipa** merupakan faktor paling dominan dalam menentukan kondisi pipeline.
- Random Forest memberikan **performa terbaik** dalam mengklasifikasikan kondisi pipeline.
- Pendekatan predictive maintenance **lebih efektif** dibandingkan maintenance konvensional.
- Pipeline dengan **degradasi tinggi** perlu diprioritaskan untuk inspeksi dan perbaikan.
- Model berpotensi dikembangkan dengan data sensor dan sistem monitoring real-time.



Thank You

PERTAMINA
KILANG PERTAMINA
INTERNASIONAL