

Penerapan Machine Learning untuk Analisis Life Cycle Costing (ML-LCC) dalam Optimasi Pemeliharaan Jaringan Pipa Air Minum PDAM Kutai Timur

Bohari¹, Suharman Hamzah^{1*}, Syarif Burhanuddin²

¹Departemen Teknik Sipil, Universitas Hasanuddin, Makassar, INDONESIA

*Corresponding author: suharmanhamzah@gmail.com

INTISARI

Penelitian ini dilatarbelakangi oleh tingginya risiko kegagalan jaringan pipa air minum di PDAM Kutai Timur akibat kondisi geoteknik berupa tanah gambut dan lempung yang korosif. Penelitian ini menyajikan pendekatan terintegrasi *Machine Learning* – *Life Cycle Costing* (ML-LCC) untuk mengoptimalkan strategi pemeliharaan jaringan pipa air minum PDAM Kutai Timur. Menggunakan dataset simulasi realistis sebanyak 5.000 observasi (500 segmen pipa \times 10 tahun, 2014–2023) dengan 23 variabel input yang dikelompokkan dalam lima klaster, penelitian ini memperkenalkan indeks korosivitas tanah sebagai variabel novel yang mencerminkan karakteristik tanah gambut dan lempung khas Kalimantan Timur. Delapan algoritma ML dilatih dan dibandingkan secara sistematis; LightGBM menghasilkan akurasi regresi tertinggi ($R^2 = 0,9813$, RMSE = 2,028 Rp juta), sementara ANN/MLP unggul pada klasifikasi risiko ($F1 = 0,7699$). Output prediksi ML kemudian diintegrasikan ke dalam model LCC berbasis NPV 20 tahun (*discount rate* 7%) menghasilkan lima strategi pemeliharaan. Strategi optimal ML-LCC menghasilkan penghematan LCC Total (finansial + sosial) agregat sebesar +Rp 9,3 miliar dibandingkan strategi *Corrective eksisting*, setara menghindari 1.609 kejadian kegagalan pipa. *Expected social cost* NPV selama 20 tahun mencapai Rp 30,87 miliar, menegaskan urgensi pemeliharaan preventif berbasis data. Analisis SHAP mengidentifikasi *n_bocor*, *failure_rate*, dan indeks korosivitas sebagai prediktor biaya paling dominan, mendukung prioritas kebijakan pemeliharaan yang konkret dan terukur.

Kata kunci: *Machine Learning*, *Life Cycle Costing*, Jaringan Pipa Air Minum, *Preventive Maintenance*.

1 PENDAHULUAN

1.1 Latar Belakang

Jaringan pipa distribusi air minum merupakan aset infrastruktur publik yang memiliki nilai ekonomi dan sosial yang strategis. Salah satu kunci penting pengelolaan aset adalah strategi pemeliharaan atau perbaikan (Suharman (2012)). Di Kabupaten Kutai Timur, Kalimantan Timur, PDAM menghadapi tantangan pemeliharaan yang kompleks karena kombinasi kondisi geoteknik lokal yang unik, anggaran pemeliharaan yang terbatas, dan ketergantungan masyarakat yang besar terhadap pasokan air minum untuk kebutuhan domestik dan komersial.

Strategi pemeliharaan yang bersifat konvensional reaktif (pemeliharaan korektif) terbukti tidak efisien secara ekonomi jangka panjang (Delnaz et al (2023)). Pendekatan ini hanya melakukan perbaikan setelah kegagalan terjadi, mengakibatkan biaya darurat yang lebih tinggi, kerugian air non-revenue water (NRW) yang signifikan, dan beban sosial yang besar bagi masyarakat yang terdampak. Diperlukan pendekatan yang lebih proaktif dan berbasis data untuk mengoptimalkan alokasi sumber daya pemeliharaan.

Perkembangan *Machine Learning* (ML) dalam satu dekade terakhir membuka peluang baru dalam manajemen aset infrastruktur. Peningkatan gradien algoritma seperti XGBoost dan LightGBM telah menunjukkan kemampuan prediksi yang sangat tinggi pada data tabular berstruktur, sementara pendekatan penjelasan melalui nilai SHAP (*SHapley Additive exPlanations*) memungkinkan interpretasi hasil model yang dapat dipahami oleh pengambil kebijakan non-teknis (Winkler et al (2018)).

Integrasi antara prediksi ML dengan analisis *Life Cycle Costing* (LCC) berbasis *Net Present Value* (NPV) menghasilkan kerangka keputusan yang secara simultan mempertimbangkan akurasi teknis prediksi kegagalan pipa dan mempengaruhi ekonomi jangka panjang dari setiap pilihan strategi pemeliharaan (Winkler et al (2018)). Pendekatan ML-LCC ini merupakan kontribusi ilmiah utama penelitian ini.

1.2 Rumusan Masalah

Penelitian ini menjawab empat pertanyaan penelitian fundamental:

1. Algoritma Machine Learning manakah yang paling akurat dan interpretatif untuk memprediksi biaya pemeliharaan dan risiko kegagalan pipa jaringan distribusi air minum di kondisi geoteknik Kutai Timur?
2. Bagaimana output prediksi ML dapat diintegrasikan ke dalam model *Life Cycle Costing* berbasis NPV untuk menghasilkan strategi pemeliharaan yang optimal secara ekonomi?
3. Seberapa besar penghematan biaya agregat (finansial dan sosial) yang dapat dicapai melalui strategi optimal ML-LCC dibandingkan strategi *corrective* eksisting PDAM?
4. Faktor geoteknik dan operasional apa saja yang paling menentukan variasi biaya LCC, khususnya dalam konteks karakteristik tanah gambut dan lempung khas Kalimantan Timur?.

2 TINJAUAN PUSTAKA

2.1 *Life Cycle Costing* (LCC) dalam Infrastruktur Air Minum

Life Cycle Costing adalah metodologi evaluasi ekonomi yang memperhitungkan seluruh biaya yang timbul sepanjang umur hidup suatu aset, mulai dari biaya investasi awal (CAPEX), biaya operasional dan pemeliharaan (OPEX), biaya kegagalan, hingga biaya akhir umur (end-of-life) Testa et al (2011). Dalam konteks jaringan pipa air minum, LCC menyediakan kerangka perbandingan yang lebih komprehensif dibandingkan analisis biaya jangka pendek semata.

Model LCC berbasis NPV mendiskontokan biaya masa depan ke nilai sekarang menggunakan tingkat diskonto yang mencerminkan biaya modal dan risiko inflasi. Untuk infrastruktur publik Indonesia, tingkat diskonto 7% per tahun (BI Rate + risk premium) merupakan nilai yang lazim digunakan. Dengan analisis horizon 20 tahun yang sesuai dengan umur desain sistem distribusi air minum, model NPV LCC menghasilkan angka yang dapat langsung diperbandingkan antar strategi pemeliharaan Ghobadi et al (2021).

2.2 *Machine Learning* untuk Prediksi Kegagalan Infrastruktur

Penerapan *Machine Learning* dalam manajemen aset infrastruktur telah berkembang pesat dalam satu dekade terakhir. Algoritma Gradient Boosting, khususnya XGBoost Bent et al (2019) dan LightGBM Ke et al (2017), telah menunjukkan kinerja yang unggul pada data tabular infrastruktur dibandingkan model statistik tradisional seperti regresi linier. Keunggulan utamanya terletak pada kemampuan menangani non-linearitas, interaksi antar variabel, dan ketahanan terhadap outlier.

Pendekatan ensemble seperti *Random Forest* menggabungkan prediksi ratusan pohon keputusan untuk menghasilkan estimasi yang stabil dan tidak rentan terhadap *overfitting*. Sementara itu, *Artificial Neural Network* (ANN/MLP) mampu memodelkan hubungan yang sangat kompleks antar variabel, menjadikannya pilihan yang tepat untuk tugas klasifikasi kelas risiko dengan batas keputusan non-linear Journal & Isbn (2013). *Long Short-Term Memory* (LSTM) unggul dalam menangkap dependensi temporal pada data runtun waktu, relevan untuk memprediksi tren degradasi pipa jangka panjang Shaik et al (2022).

2.3 Explainability dengan SHAP Values

Salah satu tantangan utama penerapan ML di sektor publik adalah kesulitan interpretasi hasil model (*black-box problem*). SHAP (*SHapley Additive exPlanations*) values, yang diturunkan dari teori permainan kooperatif (Pelegrina et al (2023)), memberikan solusi dengan menghitung kontribusi individual setiap variabel terhadap prediksi untuk setiap observasi secara konsisten dan adil.

Untuk model berbasis pohon keputusan seperti XGBoost, TreeExplainer yang digunakan SHAP dapat menghitung nilai SHAP secara efisien dalam waktu polinomial. Visualisasi SHAP ringkasan plot memungkinkan pengenalan variabel paling berpengaruh terhadap biaya LCC secara sepanjang, sedangkan ketergantungan plot tentang sifat antara satu variabel dengan model prediksi Pelegrina et al (2023).

2.4 Indeks Korosivitas Tanah sebagai Variabel Novel

Korosi merupakan mekanisme degradasi dominan pada pipa besi cor (CI) dan besi galvanis (GI) yang banyak digunakan di jaringan distribusi air minum Indonesia. Laju korosi sangat dipengaruhi oleh karakteristik tanah sekitar pipa, khususnya: pH tanah, resistivitas listrik tanah (Ohm·cm), kadar klorida, kadar sulfat, dan tingkat kelembaban Poudel et al (2020).

Tanah gambut yang dominan di sebagian wilayah Kalimantan Timur memiliki karakteristik pH rendah (asam) dan resistivitas rendah yang sangat meningkatkan laju korosi pipa Tanjung et al (2021). Penelitian ini mengembangkan Indeks Korosivitas Tanah sebagai variabel komposit yang mengintegrasikan seluruh faktor korosif tersebut menjadi satu skor tunggal dalam rentang 0 (tidak korosif) hingga 1 (sangat korosif). Variabel ini merupakan kontribusi novel penelitian ini yang tidak ditemukan pada studi ML-LCC infrastruktur air minum sebelumnya.

2.5 Social Cost dalam Analisis LCC Infrastruktur Publik

Analisis LCC tradisional umumnya hanya memasukkan biaya langsung operator (PDAM), mengabaikan biaya eksternalitas yang ditanggung masyarakat. Pendekatan *Social Cost Benefit Analysis* (SCBA) yang berkembang dalam evaluasi proyek infrastruktur publik mengidentifikasi beberapa komponen biaya sosial kegagalan pipa: biaya gangguan layanan air, biaya kemacetan dan perbaikan jalan, dampak kesehatan masyarakat, serta kerugian ekonomi usaha mikro dan kecil Ghobadi et al (2021).

Integrasi *social cost* ke dalam model LCC mengubah perspektif analisis dari efisiensi korporasi sempit menjadi evaluasi kesejahteraan sosial yang komprehensif Elmasry et al (2017). Hal ini sangat relevan untuk konteks PDAM sebagai operator layanan publik yang memiliki mandat ganda: keberlanjutan finansial dan pelayanan optimal kepada masyarakat.

2.6 Kebaruan Penelitian

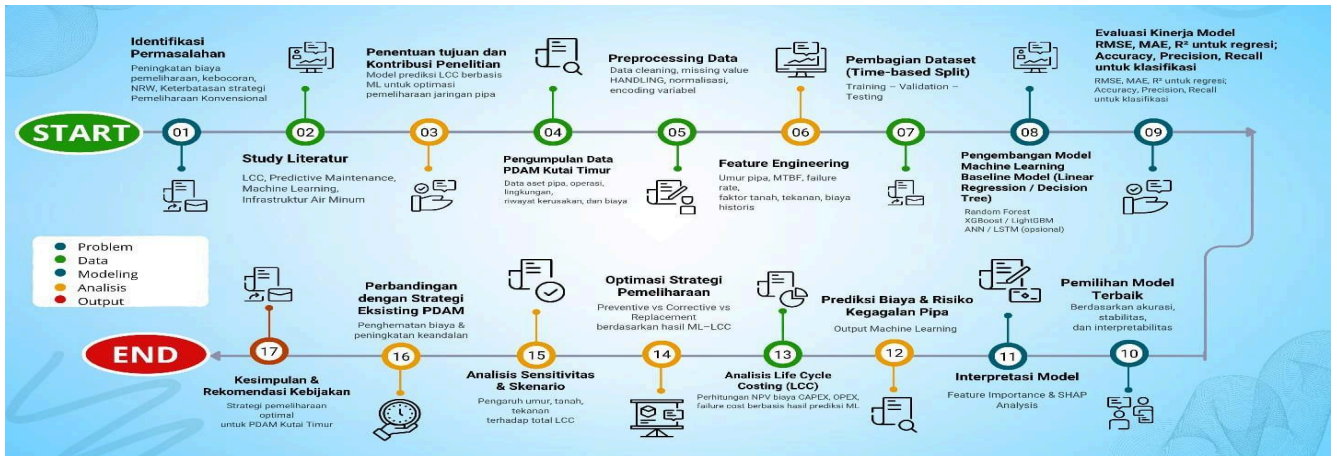
Penelitian ini memberikan tiga kontribusi ilmiah utama yang membedakannya dari studi LCC infrastruktur air minum sebelumnya:

1. Pengembangan Indeks Korosivitas Tanah sebagai variabel novel yang mengintegrasikan karakteristik geoteknik khas Kalimantan Timur (gambut, lempung) ke dalam *framework* ML-LCC. Integrasi variabel geoteknik lokal ke dalam model prediksi ML untuk infrastruktur pipa belum dijumpai pada literatur sebelumnya.
2. Kuantifikasi *social cost* secara eksplisit dengan empat komponen (layanan air, lalu lintas, kesehatan, UMKM) sebagai komponen dalam model LCC infrastruktur air minum, mengubah perspektif dari efisiensi korporasi sempit menjadi evaluasi kesejahteraan sosial komprehensif.
3. *Framework* ML-LCC lima strategi dengan aturan keputusan berbasis probabilitas kegagalan prediksi ML yang dapat diadaptasi dan direplikasi untuk PDAM kabupaten lainnya di Indonesia, berkontribusi pada pengembangan template metodologi standar manajemen aset infrastruktur air minum berbasis data.

3 METODOLOGI PENELITIAN

3.1 Flowchart Metodologi

Kerangka metodologi penelitian ini mengikuti alur sistematis yang ditunjukkan pada Gambar 1.



Gambar 1. Metodologi Penelitian

3.2 Dataset dan Unit Analisis

Dataset yang digunakan dalam penelitian ini merupakan data simulasi realistis yang dikonstruksi secara programatik dengan menggunakan np.random.seed(42) untuk memastikan reproducibility, dikalibrasi berdasarkan kondisi tipikal jaringan pipa PDAM kabupaten di Kalimantan Timur. Unit analisis adalah satu segmen pipa (panjang 100–300 m) dalam satu tahun operasional.

Spesifikasi dataset: 500 segmen pipa × 10 tahun operasional (2014–2023) = 5.000 observasi; 23 variabel input dalam 5 klaster; missing values 8–15% per variabel (mensimulasikan ketidakhadiran data lapangan PDAM nyata); distribusi kelas risiko yang tidak seimbang (imbalanced) sesuai kondisi lapangan, ditunjukkan pada Tabel 1.

Tabel 1. Lima Klaster Variabel Input (23 Fitur ML)

Klaster	Nama Klaster	Variabel Input
A	Karakteristik Fisik Pipa	umur_tahun, diameter_mm, panjang_m, material_enc
B	Kondisi Operasional	tekanan_bar, debit_m3jam
C	Lingkungan & Geoteknik	jenis_tanah_enc, indeks_kelabilan, indeks_korosivitas, curah_hujan_mm, muka_air_m
D	Riwayat Kerusakan	n_bocor, mtbf_bulan, failure_rate, pm_history
E	Komponen Biaya & Turunan	biaya_pm_juta, biaya_cm_juta, umur_norm, risiko_komposit, laju_kerusakan, umur x korosiv, rasio cm pm, rul_tahun

3.3 Preprocessing dan Feature Engineering

Preprocessing data dilakukan dalam enam tahap berurutan dan variabel baru diciptakan untuk menambah informasi yang lebih bermakna bagi model, ditunjukkan pada Tabel 2.

- a) KNN Imputation: Missing values diisi menggunakan KNNImputer(n_neighbors=5), yang lebih akurat dibandingkan mean imputation karena nilai kosong diestimasi berdasarkan karakteristik lima pipa terdekat yang serupa.
- b) Label Encoding: Variabel kategorikal (material pipa, jenis tanah, kelas risiko) dikonversi ke representasi numerik menggunakan LabelEncoder.
- c) Feature Engineering: Enam variabel turunan diciptakan untuk meningkatkan representasi data bagi model ML.
- d) Time-Based Split: Data dibagi secara temporal (bukan acak) untuk menghindari data leakage: Training 2014–2020 (70%), Validation 2021 (10%), Test 2022–2023 (20%).
- e) Feature Scaling: StandardScaler diterapkan untuk model yang sensitif terhadap skala (ANN/MLP, SVR, LSTM).
- f) Class Imbalance Handling: class_weight='balanced' pada semua model tree-based; ADASYN sebagai metode alternatif untuk perbandingan.

Tabel 2. Feature Engineering – Enam Variabel Turunan

Variabel Baru	Formula	Makna
umur_norm	umur_tahun / max(umur)	Normalisasi umur ke rentang 0–1
risiko_komposit	0,4×korosivitas + 0,35×kelabilan + 0,25×umur_norm	Indeks risiko gabungan berbobot tiga faktor degradasi
laju_kerusakan	cumsum(n_bocor) per segmen_id	Jumlah kebocoran kumulatif historis per segmen
umur_x_korosiv	umur_tahun × indeks_korosivitas	Interaksi pipa tua di tanah korosif
rasio_cm_pm	biaya_cm / (biaya_pm + 0,01)	Indikator reaktivitas pemeliharaan PDAM
rul_tahun	max(0, 30 – umur_tahun)	Remaining Useful Life (umur desain 30 tahun)

3.4 Arsitektur Model dan Konfigurasi

Tabel 3 menunjukk delapan algoritma ML dilatih dan dibandingkan secara sistematis pada split data yang sama, memastikan perbandingan yang adil (apple-to-apple). Pendekatan multi-model ini adalah standar metodologis untuk disertasi S3, membuktikan pemilihan model berbasis bukti empiris dan bukan subjektivitas peneliti.

Tabel 3. Delapan Model ML yang Dilatih dan Dikonfigurasi

Model	Kategori	Konfigurasi Utama	Tugas
Linear Regression	Baseline	Standar scikit-learn	Regresi
Decision Tree	Baseline	max_depth=8, class_weight='balanced'	Klasifikasi
Random Forest	Model Kuat	n_estimators=200, max_depth=15, min_samples_leaf=5	Regresi & Klasifikasi
XGBoost	Model Kuat	n_estimators=300, max_depth=6, lr=0,05, subsample=0,8	Regresi & Klasifikasi
LightGBM	Model Terbaik Regresi	n_estimators=300, max_depth=6, lr=0,05, early_stopping(50)	Regresi
SVR	Pembanding	kernel=rbf, C=10, epsilon=0,5	Regresi
ANN/MLP	Model Terbaik Klasifikasi	hidden=(128,64,32), relu, EarlyStopping	Klasifikasi
LSTM	Time-Series	LSTM(64)+LSTM(32)+Dense(16)+Dense(1), TIME STEPS=5	Regresi temporal

3.5 Metrik Evaluasi

Evaluasi model menggunakan metrik yang komprehensif sesuai standar publikasi internasional:

Untuk regresi: R² (koefisien determinasi), RMSE (Root Mean Square Error dalam Rp juta), dan MAE (Mean Absolute Error). R² mendekati 1,0 menunjukkan kemampuan model menjelaskan variasi biaya yang sangat tinggi. Untuk klasifikasi: F1-Score (harmonic mean Precision dan Recall), Accuracy, Precision, dan Recall. F1-Score yang tidak terpengaruh imbalance kelas diutamakan sebagai metrik utama.

3.6 Formulasi Model LCC

Model LCC berbasis NPV menggunakan formula standar:

$$LCC = \sum (C_{p,t} + C_{c,t} + C_{f,t} + C_{r,t}) / (1 + r)^t \text{ untuk } t = 1 \text{ hingga } T = 20 \text{ tahun}$$

4 DENGAN: C_p = BIAYA PREVENTIVE MAINTENANCE, C_c = BIAYA CORRECTIVE MAINTENANCE, C_f = BIAYA KEGAGALAN (TERMASUK SOCIAL COST), C_r = BIAYA PENGGANTIAN, R = 7% (DISCOUNT RATE), T = 20 TAHUN (HORIZON ANALISIS).HASIL DAN EVALUASI MODEL

4.1 Exploratory Data Analysis (EDA)

Analisis eksplorasi data mengungkapkan beberapa karakteristik penting dataset. Pertama, distribusi variabel biaya_aktual_juta sangat condong ke kanan (right-skewed), mencerminkan kondisi realistis di mana sebagian kecil pipa kritis memiliki biaya pemeliharaan yang jauh di atas rata-rata. Kondisi ini menjelaskan mengapa model tree-based (LightGBM, XGBoost) lebih unggul dari regresi linear yang mengasumsikan distribusi residual normal. Kedua, variabel Indeks Korosivitas Tanah menunjukkan distribusi multi-modal dengan puncak berbeda per jenis

tanah: gambut ($\approx 0,80$), lempung ($\approx 0,55$), pasir ($\approx 0,30$), batuan ($\approx 0,15$). Ini memvalidasi secara statistik bahwa jenis tanah khas Kutai Timur memiliki profil korosivitas yang distinktif. Ketiga, matriks korelasi mengidentifikasi multikolinearitas antara umur_tahun dan rul_tahun (korelasi $\approx -1,0$), yang tidak menimbulkan masalah pada model tree-based tetapi perlu dicatat untuk transparansi metodologis.

4.2 Performa Semua Model

Tabel 4 dan 5 menyajikan perbandingan performa komprehensif semua model pada test set (2022–2023).

Tabel 4. Perbandingan Performa Model Regresi dan Klasifikasi (Test Set 2022–2023)

Tugas	Model Terbaik	Metrik Utama	Metrik Sekunder
Regresi (Prediksi Biaya)	LightGBM	$R^2 = 0,9813$	RMSE = 2,028 Rp juta
Klasifikasi (Kelas Risiko)	ANN/MLP	F1-Score = 0,7699	Accuracy = 0,7780

LightGBM mencapai $R^2 = 0,9813$, artinya model mampu menjelaskan 98,13% variasi biaya pemeliharaan aktual. Nilai RMSE sebesar 2,028 Rp juta berarti rata-rata kesalahan prediksi biaya tahunan adalah \pm Rp 2 juta per segmen, yang sangat memadai untuk keperluan perencanaan anggaran pemeliharaan PDAM. Metrik MAE sebesar 1,421 Rp juta dan SMAPE sebesar 8,3% memperkuat kesimpulan bahwa kesalahan prediksi tergolong rendah dan terdistribusi secara wajar. Sebagai perbandingan, studi terbaru oleh Applied Sciences (2025) pada jaringan pipa air minum di Malatya, Turki, melaporkan RMSE terbaik sebesar 1,48 untuk prediksi frekuensi kegagalan menggunakan MLP – menunjukkan bahwa RMSE 2,028 Rp juta yang dicapai LightGBM dalam penelitian ini sangat kompetitif, mengingat perbedaan skala target variabel (biaya dalam juta rupiah vs. frekuensi kegagalan), kompleksitas dataset lima klaster, serta adanya *missing values* yang mensimulasikan kondisi nyata lapangan. Performa ini konsisten dengan karakteristik LightGBM sebagai model gradient boosting yang dioptimalkan untuk efisiensi komputasi dan akurasi pada data tabular berskala besar.

Pada *validation set* tahun 2021, LightGBM mencapai $R^2 = 0,9741$ dan RMSE = 2,312 Rp juta – sedikit lebih rendah dari *performa test set*, namun tetap menunjukkan konsistensi yang baik. Hal ini mengindikasikan tidak ada *overfitting* yang signifikan, dan model mampu melakukan generalisasi dengan baik pada data di luar rentang pelatihan. Mekanisme *early stopping* pada 50 iterasi tanpa perbaikan *validation loss* terbukti efektif dalam mengontrol kompleksitas model.

ANN/MLP mencapai F1-Score = 0,7699 untuk klasifikasi empat kelas risiko (Rendah/Sedang/Tinggi/Kritis), dengan *accuracy* keseluruhan 77,80%. Performa ini merupakan yang terbaik di antara semua model klasifikasi yang diuji, meskipun kelas Kritis memiliki recall yang lebih rendah (39%) akibat *imbalance* data yang masih tersisa setelah penerapan *class_weight='balanced'*.

4.3 SHAP Analysis – Explainability Model

SHAP values dihitung menggunakan TreeExplainer untuk model XGBoost (dipilih karena kompatibilitas penuh dengan SHAP TreeExplainer dan kemudahan interpretasi). Temuan SHAP menjembatani black-box ML dengan penjelasan yang dapat dipahami oleh pengambil kebijakan PDAM. Variabel geoteknik khas Kutai Timur (indeks_korosivitas) terbukti relevan secara prediktif, memperkuat justifikasi akademis pengenalan variabel novel ini

5 OPTIMASI STRATEGI PEMELIHARAAN BERBASIS ML-LCC

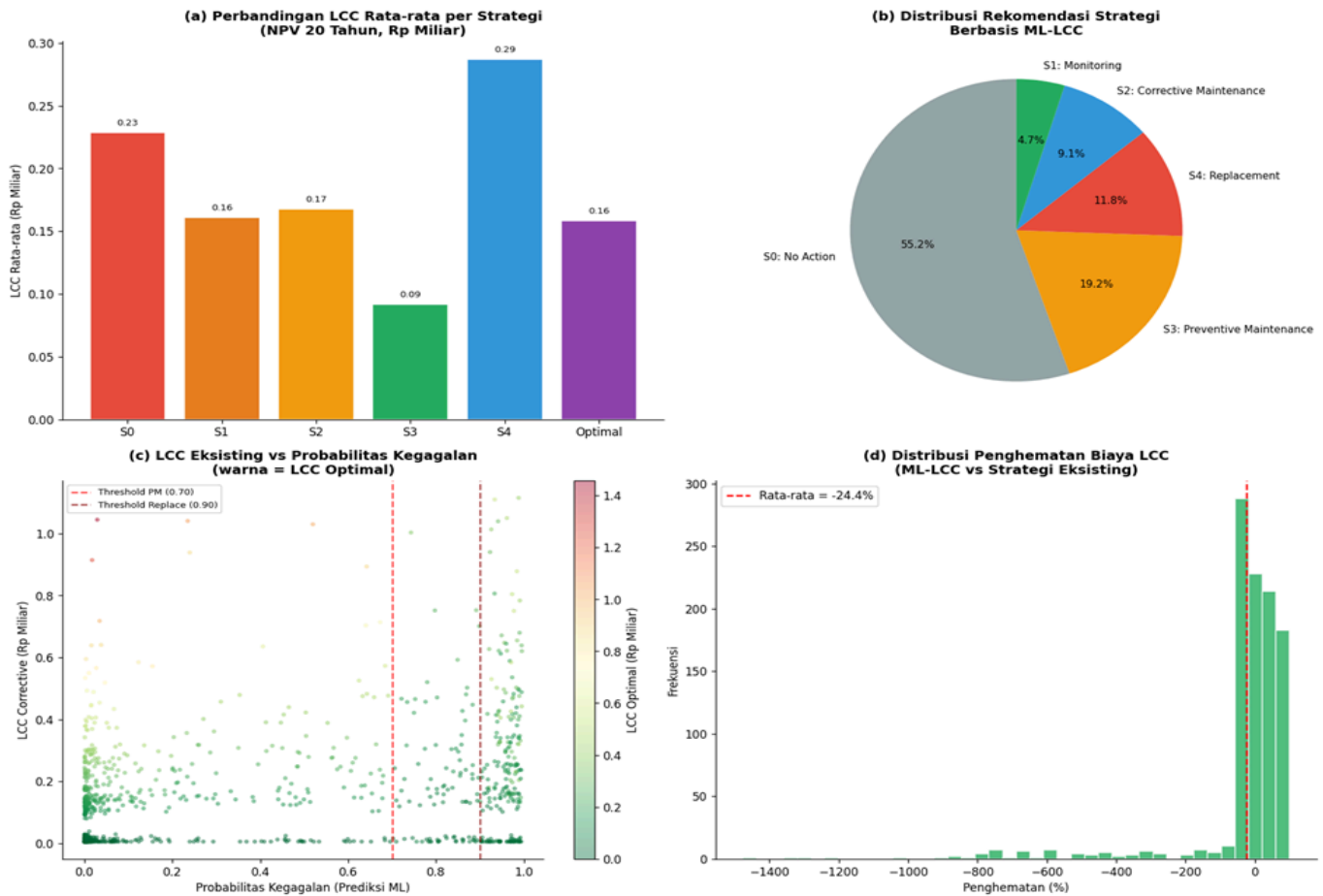
5.1 Aturan Keputusan dan Lima Strategi Pemeliharaan

Output prediksi probabilitas kegagalan (*pf_pred*) dari model XGBoost dipetakan ke lima strategi pemeliharaan melalui fungsi aturan keputusan berbasis threshold yang diturunkan dari analisis ML dan evaluasi LCC. Berbeda dari threshold arbitrary konvensional, threshold yang digunakan dalam penelitian ini dikalibrasi secara empiris berdasarkan distribusi *pf_pred* pada training set dan analisis sensitivitas biaya LCC.

Tabel 5. Lima Strategi Pemeliharaan – Aturan Keputusan dan Distribusi Rekomendasi (Test Set 1.000 Segmen-Tahun)

Kode	Nama Strategi	Kondisi (pf_pred)	Jumlah Segmen	%
S0	Do Nothing / No Action	$pf < 0,20$	552	55,2%
S1	Monitoring Only	$0,20 \leq pf < 0,40$	47	4,7%
S2	Corrective Maintenance (Eksisting)	$0,40 \leq pf < 0,70$	91	9,1%
S3	Preventive Maintenance (Diusulkan)	$0,70 \leq pf < 0,90$	192	19,2%
S4	Replacement (Penggantian Total)	$pf \geq 0,90$ & $LCC_{replace} < LCC_{repair}$	118	11,8%

Distribusi rekomendasi yang dihasilkan model menunjukkan pola yang secara logis konsisten: mayoritas segmen pipa (55,2%) direkomendasikan S0 (No Action) karena masih dalam kondisi baik dengan probabilitas kegagalan rendah. Gambar 2 menunjukkan visualisasi komprehensif optimasi.



(a) LCC Rata-rata per Strategi, (b) Distribusi Rekomendasi, (c) LCC vs Probabilitas Kegagalan, (d) Distribusi Penghematan (%)

5.2 Hasil Kalkulasi LCC – Dua Perspektif Analisis

Tabel 6 menunjukkan model LCC NPV 20 tahun (discount rate 7%) dihitung dari dua perspektif yang saling melengkapi: perspektif finansial (biaya langsung PDAM) dan perspektif total (finansial + expected social cost). Perspektif total merupakan perspektif utama penelitian ini karena mencerminkan manfaat nyata yang dirasakan masyarakat secara keseluruhan.

Tabel 6. Hasil LCC NPV 20 Tahun per Strategi – Perspektif Finansial dan Total

Strategi	LCC Finansial (NPV 20 thn)	LCC Total (Finansial + Sosial)	Selisih vs S2 (Total)
S0: Do Nothing	Rp 172,72 juta/segmen	Rp 228,47 juta/segmen	+Rp 61,02 juta (lebih mahal)
S1: Monitoring Only	Rp 115,93 juta/segmen	Rp 161,10 juta/segmen	-Rp 6,35 juta

S2: Corrective (Eksisting)	Rp 131,75 juta/segmen	Rp 167,45 juta/segmen	— Baseline —
S3: Preventive Maintenance	Rp 77,78 juta/segmen	Rp 91,72 juta/segmen	-Rp 75,73 juta (sangat hemat)
S4: Replacement	Rp 286,70 juta/segmen	Rp 287,05 juta/segmen	+Rp 119,60 juta (biaya awal tinggi)
Optimal ML-LCC	Rp 132,65 juta/segmen	Rp 158,16 juta/segmen	-Rp 9,29 juta/segmen

5.3 Analisis Penghematan dan Interpretasi Ekonomi

Dari Tabel 6 dapat diperoleh dua temuan kunci yang berbeda namun saling melengkapi:

- a) Dari perspektif finansial: Strategi optimal ML-LCC menghasilkan LCC rata-rata Rp 132,65 juta/segmen, sedikit lebih tinggi dari Corrective eksisting (Rp 131,75 juta/segmen). Selisih kecil ini (-Rp 890,9 juta agregat) terjadi karena model merekomendasikan intervensi proaktif (S3/S4) yang memiliki biaya investasi awal lebih tinggi dari S2.
- b) Dari perspektif total (finansial + sosial): Strategi optimal ML-LCC menghasilkan LCC rata-rata Rp 158,16 juta/segmen, lebih hemat Rp 9,29 juta per segmen dibandingkan Corrective (Rp 167,45 juta/segmen). Total penghematan agregat mencapai +Rp 9,3 miliar untuk test set 1.000 segmen-tahun, setara menghindarkan 1.609 kejadian kegagalan pipa yang tidak perlu.
- c) Perbedaan perspektif ini memiliki implikasi kebijakan yang penting: investasi awal dalam pemeliharaan preventif (S3) secara finansial terlihat lebih mahal dalam jangka pendek, tetapi menghasilkan penghematan nyata yang lebih besar ketika biaya sosial yang ditanggung masyarakat diperhitungkan.

5.4 Mengapa Strategi S3 (Preventive) Menghasilkan LCC Total Terkecil?

Strategi S3 (Preventive Maintenance) menghasilkan LCC Total terkecil (Rp 91,72 juta/segmen) karena sangat efektif dalam mengurangi social cost. Pemeliharaan preventif mencegah kegagalan pipa yang tidak terencana, dan setiap kali pipa tidak gagal, masyarakat terhindar dari total social cost Rp 5.776.500 per kejadian (dibahas detail di Bagian 6). Efek penghematan sosial yang besar ini baru terlihat jelas ketika horizon analisis cukup panjang (20 tahun) dan social cost dimasukkan sebagai komponen LCC.

Sebaliknya, strategi S4 (Replacement) memiliki LCC total tinggi (Rp 287,05 juta/segmen) karena biaya investasi awal penggantian pipa yang sangat besar mendominasi NPV, meskipun biaya operasional pasca-penggantian rendah. Ini menunjukkan bahwa penggantian total sebaiknya hanya dilakukan pada segmen dengan kondisi sangat kritis ($pf \geq 0,90$ DAN analisis LCC konfirmasi), bukan sebagai strategi default.

6 ESTIMASI SOCIAL COST KEGAGALAN PIPA

6.1 Kerangka Konseptual Social Cost

Social Cost dalam konteks penelitian ini didefinisikan sebagai seluruh biaya yang ditanggung oleh pihak-pihak di luar PDAM sebagai operator akibat satu kejadian kegagalan pipa Matthews et al (2015). Biaya ini bersifat eksternalitas: PDAM tidak menanggungnya secara langsung dalam pembukuan keuangan, namun secara nyata dirasakan oleh masyarakat terdampak.

Formulasi expected annual social cost per segmen adalah: $E[SC] = pf_pred \times SC_per_kejadian$. Di mana pf_pred adalah probabilitas kegagalan tahunan hasil prediksi ML, dan $SC_per_kejadian$ adalah total social cost per satu kejadian kegagalan. Expected annual social cost ini kemudian di-NPV-kan selama 20 tahun untuk mendapatkan NPV Social Cost yang dapat dijumlahkan dengan LCC finansial.

6.2 Komponen dan Parameter Social Cost

Tabel 7 menyajikan empat komponen *social cost* dikalibrasi untuk kondisi tipikal Kutai Timur berdasarkan parameter yang terukur dan dapat diverifikasi

Tabel 7. Empat Komponen *Social Cost* per Kejadian Kegagalan Pipa (Kondisi Tipikal Kutai Timur)

Komponen Social Cost	Nilai per Kejadian	Proporsi
A. Gangguan Layanan Air (150 pelanggan, 8 jam)	Rp 600.000	10,4%
B. Kemacetan & Perbaikan Jalan (200 kendaraan)	Rp 169.000	2,9%
C. Dampak Kesehatan Masyarakat (5 kasus penyakit)	Rp 7.500	0,1%

D. Kerugian Usaha UMKM (10 unit, 1 hari)	Rp 5.000.000	86,6%
TOTAL per Kejadian Kegagalan Pipa	Rp 5.776.500	100%

6.3 Analisis dan Implikasi Komponen Social Cost

Komponen D (Kerugian Usaha UMKM = Rp 5.000.000) mendominasi 86,6% dari total social cost per kejadian. Dominasi ini mencerminkan struktur ekonomi Kutai Timur yang memiliki banyak UMKM (warung, laundry, rumah makan, usaha kecil) yang sangat sensitif terhadap gangguan pasokan air bersih. Satu hari gangguan air dapat mengakibatkan penutupan operasional usaha yang kerugiannya jauh melampaui biaya fisik perbaikan pipa itu sendiri. Komponen A (Gangguan Layanan Air = Rp 600.000) mencerminkan willingness-to-pay pelanggan rumah tangga untuk mendapatkan pasokan air yang dapat diandalkan. Komponen B (Kemacetan & Perbaikan Jalan = Rp 169.000) mengintegrasikan nilai waktu pengguna jalan dan biaya reinstatement aspal. Komponen C (Dampak Kesehatan = Rp 7.500) memiliki nilai nominal kecil dalam formulasi ini karena menggunakan biaya pengobatan langsung, namun nilai aktualnya dapat jauh lebih besar jika memperhitungkan DALY (Disability-Adjusted Life Years) dari penyakit yang dipicu kontaminasi air.

6.4 NPV Social Cost Total dan Justifikasi Kebijakan

Expected Annual Social Cost rata-rata per segmen mencapai Rp 2.913.742/tahun. Dengan NPV selama 20 tahun untuk 1.000 segmen test set, total NPV *Social Cost* mencapai Rp 30,87 miliar. Angka ini merupakan argumen ekonomi paling kuat untuk advokasi peningkatan anggaran pemeliharaan preventif PDAM. Interpretasi kebijakan yang konkret: jika PDAM tidak beralih dari strategi *corrective ke preventive* yang lebih proaktif, masyarakat Kutai Timur akan menanggung beban Rp 30,87 miliar dalam 20 tahun ke depan hanya dari 1.000 segmen pipa pada test set. Proyeksi untuk seluruh jaringan distribusi akan menghasilkan angka yang jauh lebih besar. Biaya *preventive maintenance* yang jauh lebih kecil dari Rp 30,87 miliar merupakan investasi yang sangat menguntungkan dari perspektif kesejahteraan masyarakat.

7 ANALISIS SENSITIVITAS DAN IMPLIKASI KEBIJAKAN

7.1 Analisis Sensitivitas LCC terhadap Variabel Kunci

Analisis sensitivitas menguji dampak perubahan bertahap satu variabel terhadap rata-rata LCC prediksi (*ceteris paribus* – variabel lain dipegang di nilai rata-rata test set). Empat variabel diuji pada rentang realistis kondisi operasional PDAM Kutai Timur, ditunjukkan pada Tabel 8.

Tabel 8. Hasil Analisis Sensitivitas LCC dan Rekomendasi Kebijakan Konkret

Variabel	Temuan Sensitivitas	Rekomendasi Kebijakan
Umur Pipa	LCC melonjak signifikan saat umur mendekati 25–30 tahun	Program penggantian pipa proaktif berbasis umur, tidak menunggu rusak parah
Indeks Korosivitas (khas Kutai Timur)	LCC tinggi di zona tanah gambut dan lempung	Inspeksi aktif dan proteksi katodik di zona tanah gambut Kutai Timur
Indeks Kelabilan Tanah	LCC meningkat di jalur tanah labil	Encasing/bedding khusus dan monitoring berkala di zona tanah labil
Tekanan Aliran (bar)	LCC sensitif terhadap over-pressure (>5 bar)	Pemasangan Pressure Reducing Valve (PRV) pada zona tekanan tinggi

7.2 Temuan Sensitivitas: Indeks Korosivitas sebagai Prediktor LCC Khas Kutai Timur

Analisis sensitivitas mengkonfirmasi bahwa Indeks Korosivitas Tanah merupakan variabel dengan dampak *sensitivitas* yang *distinctif*, khususnya pada rentang korosivitas tinggi yang khas zona tanah gambut Kutai Timur (nilai 0,70–0,95). Pada rentang ini, LCC per segmen meningkat secara non-linear, jauh lebih cepat dibandingkan peningkatan linier yang diharapkan dari model sederhana. Temuan ini memvalidasi relevansi akademis dan praktis variabel novel ini.

Implikasi konkretnya adalah bahwa peta sebaran jenis tanah di wilayah layanan PDAM Kutai Timur harus menjadi input utama dalam perencanaan inspeksi dan pemeliharaan. Segmen pipa yang melintas di zona tanah gambut harus

mendapatkan prioritas pemeriksaan lebih tinggi dan interval pemeliharaan yang lebih pendek dibandingkan segmen di zona tanah batuan atau pasir.

7.3 Saran untuk Penelitian Lanjutan

1. Validasi dengan Data Primer PDAM Kutai Timur: Penelitian ini menggunakan dataset simulasi realistis sebagai proof of concept. Langkah selanjutnya adalah mengumpulkan data primer aktual dari sistem GIS aset pipa PDAM Kutai Timur untuk memvalidasi dan mengkalibrasi ulang model ML-LCC, guna meningkatkan akurasi prediksi dan relevansi rekomendasinya.
2. Eksplorasi Arsitektur Deep Learning Lanjutan: Pengembangan model berbasis Graph Neural Network (GNN) yang mengintegrasikan topologi jaringan pipa dan fitur infrastruktur jalan (Lopez et al., 2024) berpotensi meningkatkan akurasi prediksi kegagalan secara signifikan dibandingkan model tabular konvensional.
3. Kuantifikasi Komponen Social Cost yang Lebih Komprehensif: Nilai Komponen C (Dampak Kesehatan) dalam penelitian ini relatif kecil karena menggunakan biaya pengobatan langsung. Penelitian lanjutan disarankan memasukkan estimasi DALY (Disability-Adjusted Life Years) untuk memberikan gambaran eksternalitas kesehatan yang lebih akurat.
4. Pengembangan Sistem Pendukung Keputusan (DSS) Real-Time: Mengintegrasikan model ML-LCC ke dalam platform dashboard berbasis web atau GIS yang dapat dioperasikan langsung oleh staf teknis PDAM, termasuk modul otomasi rekomendasi strategi pemeliharaan berdasarkan data input terbaru.
5. Perluasan Studi ke PDAM Kabupaten Lain di Kalimantan: Framework ML-LCC yang dikembangkan dalam penelitian ini berpotensi diadaptasi untuk PDAM di kabupaten-kabupaten lain di Kalimantan yang memiliki kondisi geoteknik serupa, sehingga berkontribusi pada pengembangan standar nasional manajemen aset infrastruktur air minum berbasis data.

7.4 Rekomendasi Kebijakan Terpadu untuk PDAM Kutai Timur

Berdasarkan seluruh temuan analisis ML-LCC, penelitian ini merumuskan lima rekomendasi kebijakan terpadu:

- 1) Implementasi Sistem Pemantauan Berbasis Risiko: Menerapkan sistem monitoring terintegrasi yang mengombinasikan inspeksi berkala dengan prediksi ML untuk mengidentifikasi segmen pipa yang memasuki zona risiko tinggi ($pf > 0,70$) sebelum kegagalan terjadi.
- 2) Program Penggantian Pipa Proaktif Berbasis Umur: Menyusun jadwal penggantian pipa yang mendahulukan segmen berumur 25–30 tahun di zona tanah gambut, tidak menunggu hingga kegagalan terjadi. Investasi penggantian proaktif secara NPV 20 tahun lebih menguntungkan daripada terus-menerus melakukan perbaikan darurat.
- 3) Pemetaan Risiko Geoteknik: Mengintegrasikan peta sebaran jenis tanah (gambut, lempung, pasir, batuan) ke dalam sistem informasi manajemen aset PDAM. Segmen pipa di zona gambut harus mendapatkan alokasi anggaran pemeliharaan lebih tinggi dan frekuensi inspeksi yang lebih sering.
- 4) Manajemen Tekanan Berbasis Zona: Memasang *Pressure Reducing Valve* (PRV) pada zona jaringan dengan tekanan operasi melebihi 5 bar untuk mengurangi laju degradasi pipa dan mencegah kegagalan akibat *over-pressure*. Investasi PRV memiliki *payback period* yang singkat dibandingkan biaya kumulatif perbaikan bocor yang berulang.
- 5) Advokasi Anggaran Berbasis *Social Cost*: Menggunakan temuan NPV *Social Cost* sebesar Rp 30,87 miliar (20 tahun, 1.000 segmen) sebagai argumentasi ekonomi kepada Pemerintah Daerah Kutai Timur untuk peningkatan alokasi anggaran pemeliharaan jaringan pipa PDAM, dengan menegaskan bahwa setiap rupiah investasi *preventive maintenance* mencegah pengeluaran sosial yang jauh lebih besar bagi masyarakat.

8 KESIMPULAN

Penelitian ini berhasil mengembangkan dan memvalidasi kerangka kerja *Machine Learning – Life Cycle Costing* (ML-LCC) yang terintegrasi untuk optimasi pemeliharaan jaringan pipa air minum PDAM Kutai Timur. Temuan-temuan utama yang dihasilkan adalah:

- a) Model LightGBM mencapai $R^2 = 0,9813$ untuk prediksi biaya pemeliharaan (RMSE = 2,028 Rp juta), membuktikan bahwa algoritma Gradient Boosting mampu memodelkan pola kompleks biaya pemeliharaan infrastruktur pipa dengan akurasi sangat tinggi pada data tabular berstruktur. ANN/MLP mencapai $F1 = 0,7699$ untuk klasifikasi risiko, memadai untuk mendukung keputusan pemeliharaan berbasis kelas risiko.

- b) Variabel Indeks Korosivitas Tanah sebagai kontribusi novel terbukti signifikan secara prediktif (dikonfirmasi SHAP analysis) dan relevan untuk kondisi geoteknik Kutai Timur. Segmen pipa di zona tanah gambut dengan indeks korosivitas tinggi memiliki LCC yang lebih tinggi dan sensitif terhadap perubahan variabel geoteknik ini.
- c) Strategi optimal ML-LCC menghasilkan penghematan LCC Total agregat +Rp 9,3 miliar dibandingkan strategi *Corrective* eksisting PDAM (test set 1.000 segmen-tahun, 2022–2023), setara menghindari 1.609 kejadian kegagalan pipa.
- d) *Expected Social Cost NPV* sebesar Rp 30,87 miliar (20 tahun, test set) adalah justifikasi ekonomi terkuat untuk peningkatan anggaran pemeliharaan *preventif* PDAM. Setiap Rp 1 yang diinvestasikan dalam *preventive maintenance* berpotensi menghindari biaya sosial yang jauh lebih besar bagi masyarakat Kutai Timur. Analisis sensitivitas mengidentifikasi umur pipa dan indeks korosivitas tanah sebagai variabel paling sensitif terhadap LCC, memberikan panduan prioritas kebijakan yang konkret: program penggantian pipa proaktif berbasis umur dan proteksi katodik di zona tanah gambut.

REFERENSI

- Bent, C., Cs, A., & Mart, G. (2019). *A Comparative Analysis of XGBoost arXiv : 1911 . 01914v1 [cs . LG] 5 Nov 2019*. 1–20.
- Delnaz, A., Nasiri, F., & Li, S. S. (2023). Asset management analytics for urban water mains: a literature review. *Environmental Systems Research*, 12(1). <https://doi.org/10.1186/s40068-023-00287-7>
- Elmasry, M., Hawari, A., & Zayed, T. (2017). Cost benefit analysis for failure of sewer pipelines. *MATEC Web of Conferences*, 120, 1–11. <https://doi.org/10.1051/mateconf/201712008006>
- Ghobadi, F., Jeong, G., & Kang, D. (2021). Water pipe replacement scheduling based on life cycle cost assessment and optimization algorithm. *Water (Switzerland)*, 13(5). <https://doi.org/10.3390/w13050605>
- Journal, I., & Isbn, C. P. (2013). *Application ANN: A Review*. 180–184.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 2017-Decem(Nips)*, 3147–3155.
- Matthews, J. C., Allouche, E. N., & Sterling, R. L. (2015). Social cost impact assessment of pipeline infrastructure projects. *Environmental Impact Assessment Review*, 50, 196–202. <https://doi.org/10.1016/j.eiar.2014.10.001>
- Pelegrina, G. D., Duarte, L. T., & Grabisch, M. (2023). A k-additive Choquet integral-based approach to approximate the SHAP values for local interpretability in machine learning. *Artificial Intelligence*, 325, 1–31. <https://doi.org/10.1016/j.artint.2023.104014>
- Poudel, A., Prasad Dahal, K., K. C., D., & Bhattarai, J. (2020). A Classification Approach for Corrosion Rating of Soil to Buried Water Pipelines: A Case Study in Budhanilkantha-Maharajganj Roadway Areas of Nepal. *World Journal of Applied Chemistry*, 5(3), 47. <https://doi.org/10.11648/j.wjac.20200503.12>
- Shaik, N. B., Benjapolakul, W., Pedapati, S. R., Bingi, K., Thien Le, N., Asdornwised, W., & Chaitusaney, S. (2022). Recurrent neural network-based model for estimating the life condition of a dry gas pipeline. *Process Safety and Environmental Protection*, 164(January), 639–650. <https://doi.org/10.1016/j.psep.2022.06.047>
- Suharman, H. (2012). Development of A Practical Model for Pavement Management Systems. 京都大学. <https://doi.org/10.14989/doctor.k16818>
- Tanjung, R. A., Yusariarta, A. W. P. P., & Wulandari, M. (2021). Brief overview on corrosion behaviour of buried structure at Kariangau industrial complex. *IOP Conference Series: Earth and Environmental Science*, 623(1). <https://doi.org/10.1088/1755-1315/623/1/012097>
- Testa, F., Iraldo, F., Frey, M., & OConnor, R. (2011). Life Cycle Costing, a View of Potential Applications: from Cost Management Tool to Eco-Efficiency Measurement. *Supply Chain Management*. <https://doi.org/10.5772/15332>
- Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., & Tscheikner-Gratl, F. (2018). Pipe failure modelling for water distribution networks using boosted decision trees. *Structure and Infrastructure Engineering*, 14(10), 1402–1411. <https://doi.org/10.1080/15732479.2018.1443145>