

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR ISTILAH	xiv
INTISARI.....	xix
<i>ABSTRACT</i>	xx
BAB I PENDAHULUAN	
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	6
1.3. Batasan Masalah.....	7
1.4. Tujuan Penelitian.....	8
1.5. Manfaat Penelitian.....	9
BAB II TINJAUAN PUSTAKA	
2.1. Tinjauan Pustaka	10
2.2. Keaslian Penelitian	12
2.3. Landasan Teori.....	17
BAB III METODE PENELITIAN	
3.1. Jenis, Sifat dan Pendekatan Penelitian	39
3.2. Metode Pengumpulan Data	40
3.3. Metode Analisis Data	42
3.4. Alur Penelitian.....	46

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

4.1. Hasil Penelitian	53
4.2. Pembahasan.....	72

BAB V PENUTUP

5.1. Kesimpulan.....	77
5.2. Saran.....	78

DAFTAR PUSTAKA

LAMPIRAN



DAFTAR TABEL

Tabel 2.1. Matriks Literatur Review dan Posisi Penelitian.....	14
Tabel 4.1. Pembagian Kelompok Stratum Berdasarkan Kuantil	54
Tabel 4.2. Komparasi Waktu Komputasi XGBoost dan CatBoost	55
Tabel 4.3. Hasil Evaluasi <i>Baseline</i> XGBoost.....	56
Tabel 4.4. Hasil Evaluasi <i>Baseline</i> CatBoost.....	56
Tabel 4.5. Konfigurasi Hiperparameter Terbaik XGBoost-Optuna.....	58
Tabel 4.6. Konfigurasi Hiperparameter Terbaik CatBoost-Optuna.....	59
Tabel 4.7. Komparasi Kinerja Model (Baseline, Random Search, dan Optuna....	60
Tabel 4.9. Komparasi Metrik Evaluasi Model Hasil Optuna.....	66
Tabel 4.10. Hasil Ekstraksi Anomali Audit (Ambang Batas Deviasi 10%)	67
Tabel 4.11. Contoh Data Terindikasi Potensi Anomali (Model XGBoost)	70
Tabel 4.12. Contoh Data Terindikasi Potensi Anomali (Model CatBoost)	71

DAFTAR GAMBAR

Gambar 3.1. Diagram Alir Pengumpulan Data	41
Gambar 3.2. Diagram Alir Penelitian	47
Gambar 4.1. Komparasi WAPE : XGBoost dan CatBoost	62
Gambar 4.2. Sebaran Rasio Penyusutan dan Aktual XGBoost-Optuna.....	63
Gambar 4.3. Feature Importance Model XGBoost-Optuna	65
Gambar 4.4. Deteksi Anomali pada Model XGBoost	68
Gambar 4.5. Deteksi Anomali pada Model CatBoost.....	68
Gambar 4.6. Distribusi Error Model XGBoost	69
Gambar 4.7. Distribusi Error Model CatBoost	69

DAFTAR ISTILAH

Berikut adalah daftar istilah teknis dan akademis yang telah diurutkan berdasarkan abjad beserta penomorannya:

1. Akumulasi Penyusutan : Total akumulatif dari nilai penyusutan suatu aset yang dihitung sejak awal tahun perolehan hingga tahun pelaporan akuntansi berjalan.
2. Barang Milik Daerah (BMD) : Seluruh aset, kekayaan, atau barang yang dibeli maupun diperoleh atas beban Anggaran Pendapatan dan Belanja Daerah (APBD) atau berasal dari perolehan lainnya yang sah.
3. *Baseline* (Model Dasar) : Model algoritme awal yang dilatih secara independen menggunakan konfigurasi hiperparameter bawaan (*default*) untuk menetapkan tolok ukur (*benchmark*) kinerja sebelum optimasi lanjutan diterapkan.
4. CatBoost : Algoritme *ensemble learning* berbasis *gradient boosting* yang memiliki arsitektur bawaan yang dioptimalkan untuk memproses fitur-fitur kategorikal secara efisien tanpa memerlukan rekayasa dimensi yang ekstensif.
5. *Data Leakage* (Kebocoran Data) : Kondisi cacat komputasi di mana model *machine learning* mempelajari informasi redundan (seperti data ganda) atau data uji, sehingga menyebabkan bias dan tingkat akurasi yang tidak valid.
6. De-duplikasi : Tahapan prapemrosesan data yang bertujuan untuk mendeteksi, mengidentifikasi, dan mengeliminasi rekaman data redundan atau ganda guna menghasilkan observasi berprofil unik.

7. *Ensemble Learning* : Paradigma pemodelan *machine learning* yang menggabungkan beberapa algoritme (seperti pohon keputusan) untuk menghasilkan satu model prediktif yang jauh lebih akurat dan stabil dibandingkan model tunggal.
8. *Feature Importance* (Tingkat Kepentingan Fitur) : Metrik interpretasi struktural yang merepresentasikan seberapa besar tingkat kontribusi atau pembobotan suatu variabel prediktor dalam memecah cabang pohon keputusan pada model.
9. *Gradient Boosting Decision Tree* (GBDT) : Arsitektur *machine learning* yang membangun model ansambel berupa pohon keputusan secara sekuensial (bertahap), di mana setiap pohon baru dilatih secara spesifik untuk mengoreksi galat (residu) dari agregasi pohon-pohon sebelumnya.
10. Hiperparameter : Parameter struktural pada arsitektur *machine learning* yang nilainya harus diatur secara manual sebelum proses pelatihan dimulai guna mengontrol kompleksitas pembelajaran dan mencegah galat.
11. *Knowledge Discovery in Databases* (KDD) : Kerangka kerja sistematis dan terstruktur yang mencakup tahapan pemahaman data, prapemrosesan, transformasi, pemodelan data *mining*, hingga evaluasi untuk mengekstraksi informasi berharga.
12. Koefisien Determinasi (R^2) : Metrik statistik yang menunjukkan persentase seberapa baik variabel-variabel prediktor secara proporsional mampu menjelaskan variabilitas atau pergerakan variabel target.

13. *Mean Absolute Error (MAE)* : Metrik evaluasi regresi statistik yang mengukur rata-rata dari selisih nilai absolut antara angka prediksi komputasi dengan angka aktual di lapangan.
14. Optimasi *Bayesian* : Pendekatan matematis probabilistik yang digunakan untuk mencari konfigurasi hiperparameter algoritme dengan memodelkan rekam jejak riwayat pencarian sebelumnya guna menavigasi titik konvergensi yang lebih cerdas.
15. *Optuna* : Kerangka kerja perangkat lunak untuk optimasi hiperparameter otomatis yang ditenagai oleh algoritme pencarian probabilistik seperti *Tree-structured Parzen Estimator (TPE)*.
16. *Ordered Target Encoding* : Mekanisme internal pada arsitektur CatBoost yang dirancang khusus untuk memetakan hierarki atribut kategorikal guna meredam kebocoran target dan mengelola disparitas variabel numerik.
17. *Overfitting (Kelebihan Kecocokan)* : Fenomena di mana model *machine learning* menghafal rincian atau *noise* dari dataset latih secara berlebihan, sehingga akurasi anjlok secara drastis saat diuji pada data baru.
18. Potensi Anomali (*red flag*) : Label indikator yang diberikan secara otomatis oleh sistem komputasional kepada entri aset yang memiliki deviasi prediksi penyusutan melampaui ambang batas materialitas audit (misal: di atas 10%).
19. *Random Search* : Metode pendekatan heuristik yang mengkalibrasi hiperparameter dengan menguji berbagai kombinasi nilai secara acak tanpa panduan riwayat probabilitas sebelumnya.

20. SIMBADAMAS : Sistem Informasi Barang Daerah milik Pemerintah Kabupaten Banyumas yang berfungsi sebagai basis data utama operasional pencatatan aset dan sumber ekstraksi data mentah penelitian.
21. Sistem Pendukung Keputusan (*Decision Support System*) : Sistem komputasional yang dirancang untuk memfasilitasi dan mengarahkan pengguna (seperti auditor) dalam mengambil keputusan berbasis data, mengubah verifikasi manual menjadi verifikasi tepat sasaran.
22. Stratifikasi Kuantil Berfrekuensi Setara (*Equal-Frequency Quantile Stratification*) : Teknik partisi data yang mengelompokkan observasi ke dalam lapisan (*stratum*) secara berurutan di mana setiap lapisan memiliki distribusi frekuensi/jumlah sampel yang relatif identik.
23. *Target Leakage* (Kebocoran Target) : Kelemahan pada arsitektur algoritme *boosting* konvensional di mana variabel yang seharusnya ditebak (*target*) malah mempengaruhi nilai dari variabel prediktor selama proses komputasi berlangsung.
24. *Tree-structured Parzen Estimator* (TPE) : Algoritme pemodelan probabilistik yang menjadi mesin penggerak utama di dalam kerangka kerja *Optuna*. Algoritme ini bekerja dengan membangun model probabilitas dari riwayat evaluasi sebelumnya guna menavigasi pencarian hiperparameter yang paling menjanjikan secara cerdas.
25. *Underfitting* (Bias Prediksi) : Kondisi di mana suatu algoritme prediksi terlalu sederhana atau terlalu membatasi diri sehingga gagal menangkap atau merepresentasikan pola esensial dari data aktual.

26. Varians Nominal Ekstrem : Karakteristik sekumpulan data aset yang memiliki disparitas (rentang selisih) rentang total perolehan finansial yang sangat luas (berkisar dari ratusan ribu hingga miliaran rupiah).
27. *Weighted Absolute Percentage Error* (WAPE) : Metrik evaluasi galat persentase yang diagregasi dan dibobotkan, yang sangat andal digunakan untuk mengukur kinerja algoritme pada distribusi data moneter yang bernilai heterogen/fluktuatif.
28. XGBoost : Singkatan dari *Extreme Gradient Boosting*, yakni arsitektur *machine learning* populer berbasis pohon keputusan ansambel yang dirancang sangat optimal untuk meningkatkan kecepatan proses dan presisi prediksi matematis.