Socioeconomic Occupation Proxies, Recorded Legal Representation, and Prosecutorial Demand Anchoring in Low-Value Theft Sentencing: An Observational Content Analysis of the Indo-Law Corpus across Five Provinces in Java (2005–2019)

Sandro Louise Oliveirra, PhD[1], Henry Peter Paul, PhD[2], Harun Urrashid, PhD[3], Dais Susilo, SH, MH[3*], Ervin Munandar, PhD[4], Christian Napitupulu, SH, MH[5], Ahmad Erza, SH, MH[3], Dr Yuniarti Maretha Pasaribu, S.Sn[3], Firzan Dahlan, S.Li., PhD[3], Anita Havyasari, S.Par, PhD[6]

[1]ESBA Institute of Tourism, Belo, Brazil

[2]Miami History Museum, Miami, United States

[3]Department of Social Sciences, Enigma Institute, Palembang, Indonesia

[4]Buana Legal Institute, Banyuwangi, Indonesia

[5]Sanskrit Institute, Jakarta, Indonesia

[6]Nusantara Tourism and Heritage School, Makassar, Indonesia

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.101000689

Received: 21 October 2026; Accepted: 26 October 2026; Published: 16 November 2026

ABSTRACT

Background: Unequal access to legal counsel and socioeconomic vulnerability may shape criminal sentencing, yet quantitative evidence from Indonesian court decisions remains scarce.

Objective: This study evaluated whether defendant occupation and recorded legal representation are independently associated with sentence length in low-value theft decisions after accounting for prosecutorial demands and offense characteristics.

Methods: All 22,630 XML criminal decisions in Indo-Law (commit 67340338adbc63021630947d905627eef0d7e43f; Java, 2005–2019) were screened. Theft decisions (n = 2,398) with readable loss (n = 2,138) ≤ IDR 2,500,000, zero material optical character recognition conflicts, and no above-threshold phrases formed the primary cohort (N = 148). Mann–Whitney U, Fisher's exact, ordinary least squares on log sentence days with HC3 standard errors (n = 120), and logistic regression (> 540 days) were estimated.

Results: Across N = 148 decisions, median sentence was 540 days (interquartile range [IQR]: 360–810; range: 0–4,320); 146 (98.6%) exceeded the 90-day benchmark (0/148 cited Article 364) and 133 (89.9%) involved aggravation or violence (Articles 363/365). Counsel was assisted in 19 (12.8%; median 300 days, IQR: 165–585), unassisted in 70 (47.3%; median 600 days, IQR: 375–787), and unrecorded in 59 (39.9%; median 630 days, IQR: 450–900). Crude comparison favored assisted defendants (p = 0.019), but multivariable adjustment (n = 120; R^2 = 0.739) attenuated the counsel difference to −8.9% (95% confidence interval [CI]: −29.5% to +17.8%, p = 0.477; logistic adjusted odds ratio = 0.85, 95% CI: 0.18–3.89, p = 0.831). Occupation showed no independent association, whereas prosecutorial demand (median 720 days) strongly predicted sentence length (β = 0.832, 95% CI: 0.702–0.963, p < 0.001; +8.3% per +10% demand).

Conclusions: Crude sentencing differences by recorded counsel reflect confounding and mediation by prosecutorial demand and aggravated offense composition rather than independent causal disparities.

Keywords: *Indonesia; Indo-Law; legal aid; legal natural language processing; prosecutorial demand; sentencing disparity; theft*

INTRODUCTION

Judicial decision documents preserve an authoritative institutional trace of criminal adjudication, recording the defendant's demographic identity, procedural history of legal representation, prosecutorial indictment and sentencing demand, evidentiary findings, judicial considerations of aggravating and mitigating circumstances, and the final dispositive verdict.[1,2] In computational legal studies and legal natural language processing (NLP), large-scale digital repositories of court judgments enable empirical legal scholars to investigate structural determinants of sentencing outcomes without manually retrieving thousands of unindexed paper or portable document format (PDF) files.[1–3] Within the Indonesian civil-law jurisdiction, the open-source Indo-Law corpus—compiled by the Information Retrieval and Natural Language Processing Laboratory at Universitas Indonesia—provides 22,630 first-instance general criminal court decisions (inkracht) harvested from the Supreme Court Decision Directory across five provinces on the island of Java (West Java, Central Java, East Java, DKI Jakarta, and DI Yogyakarta) and annotated into eleven standardized structural XML sections.[1,2] This structured corpus architecture creates an empirical foundation for examining how legal and extralegal factors interact during routine criminal sentencing in Indonesian district courts.[1,3]

Rigorous quantitative evaluation of socioeconomic disparities and legal aid access in Indonesian property crime sentencing requires resolving three foundational doctrinal and methodological issues.[4–10] First, in Indonesian criminal law, the empirical category of nominally low-value theft defined by monetary loss is not legally synonymous with the statutory category of minor theft (tindak pidana pencurian ringan).[4,11–15] Under the legacy Indonesian Criminal Code (Kitab Undang-Undang Hukum Pidana [KUHP], derived from the Dutch Wetboek van Strafrecht), ordinary theft is governed by Article 362 (maximum five years imprisonment), theft under aggravating circumstances (pencurian dengan pemberatan—such as offenses committed at night in an inhabited dwelling, by two or more persons acting in concert, or by breaking and entering) is governed by Article 363 (maximum seven to nine years imprisonment), minor theft is governed by Article 364 (maximum three months [90 days] imprisonment), and theft accompanied by violence or threats of violence (pencurian dengan kekerasan) is governed by Article 365 (maximum nine to twelve years imprisonment before capital/life enhancements).[4,11–15] In 2012, the Supreme Court enacted Supreme Court Regulation (Peraturan Mahkamah Agung [Perma]) Number 2 of 2012 concerning the Adjustment of the Limitations on Minor Offenses and the Amount of Fines in the Criminal Code, raising the statutory monetary threshold for Article 364 minor theft from IDR 250 to IDR 2,500,000 in response to currency depreciation.[4,11–14] Crucially, however, Article 364 explicitly excludes thefts committed in an inhabited dwelling or enclosed yard or accompanied by violent or certain aggravating modalities, and summary minor-offense proceedings (acara pemeriksaan cepat / Pid.C) are docketed separately from ordinary collegiate criminal trials (acara pemeriksaan biasa / Pid.B).[4,14,15] Consequently, a theft decision in a general criminal archive involving an explicit monetary loss of IDR 2,500,000 or less is legally prosecuted under Articles 362, 363, or 365 whenever disqualifying statutory elements are charged, meaning that the 90-day (three-month) ceiling of Article 364 serves strictly as an empirical diagnostic benchmark rather than a universally binding statutory limit for all nominally low-value thefts.[4,14,15]

Second, historical Indonesian court corpora must be situated within national penal reform.[5,10,16–18] Indonesia enacted Law Number 1 of 2023 concerning the National Criminal Code (Undang-Undang Nomor 1 Tahun 2023 tentang Kitab Undang-Undang Hukum Pidana), which entered into force on 2 January 2026 to replace the colonial-era criminal code and introduce codified sentencing guidelines, non-custodial alternatives such as social work punishment (pidana kerja sosial), and restorative justice mechanisms.[5,10,16–18] Because all 22,630 decisions in the Indo-Law repository originate from 2005 through 2019, empirical analysis of this corpus represents a historical baseline investigation of pre-reform judicial practice under the legacy KUHP and Perma Number 2 of 2012 rather than an evaluation of post-2026 statutory classifications.[1,4,5]

Third, within the sociology of punishment and focal concerns theory, courtroom sentencing reflects the interplay among offense blameworthiness, perceived community dangerousness, practical organizational constraints, defense counsel resources, and institutional alignment between prosecutorial sentence recommendations and judicial verdicts.[6–10,19–25] Empirical studies across jurisdictions show that while unadjusted comparisons frequently suggest substantial sentencing penalties against economically marginalized or unrepresented defendants, such crude gradients are heavily confounded by charge severity, recidivism, and post-indictment prosecutorial sentencing demands.[6–10,19–23] Moreover, automated extraction of monetary losses, legal representation status, and occupational categories from digitized court decisions is susceptible to optical character recognition (OCR) corruption, administrative non-recording, and unmeasured confounding, requiring deterministic validation and transparent reporting aligned with the REporting of studies Conducted using Observational Routinely-collected Data (RECORD) statement.[24–26]

The present study constructs a reproducible, RECORD-compliant lexical-numeric extraction and econometric auditing pipeline across the entire 22,630-document Indo-Law XML corpus to evaluate whether defendant occupational proxies and recorded legal representation are independently associated with custodial sentence length in nominally low-value theft decisions (explicit loss ≤ IDR 2,500,000) after accounting for prosecutorial demands and statutory offense characteristics.[1,4,25,26] Specifically, this study: (1) constructs and manually validates a primary low-value theft analytic cohort (N = 148) and a conservative cumulative-loss sensitivity cohort (n = 131) free of material OCR digit–word conflicts; (2) quantifies bivariate distributions of custodial sentence length (in days), prosecutorial demands, and > 90-day diagnostic benchmark exceedance across recorded legal representation status (assisted, unassisted, and unrecorded) and seven lexical defendant occupation groups; (3) characterizes mitigating and aggravating judicial considerations among decisions with segmented considerations sections (n = 80) alongside formal missing-data comparisons; (4) estimates multivariable ordinary least squares (OLS) regression models on natural log-transformed sentence length with MacKinnon–White HC3 heteroskedasticity-consistent standard errors (n = 120 complete cases) and secondary multivariable logistic regression models for sentences exceeding the cohort median (> 540 days); and (5) audits all 22,630 XML files for explicit references to Article 364 KUHP and Perma Number 2 of 2012.[1,2,4,5,19–26]

METHODS

Study Design, Data Source, and Reproducible Computational Environment

This study was executed as a retrospective, observational quantitative content analysis of publicly archived secondary judicial data, reported in accordance with the RECORD statement for observational studies using routinely collected administrative and archival records.[26] The primary data source was the open-source Indo-Law repository maintained by the Information Retrieval and Natural Language Processing Laboratory at the Faculty of Computer Science, Universitas Indonesia (repository identifier: ir-nlp-csui/indo-law), locked deterministically at Git commit hash 67340338adbc63021630947d905627eef0d7e43f.[1,2] The Indo-Law corpus comprises 22,630 first-instance general criminal court decision documents (putusan pengadilan negeri tingkat pertama yang telah berkekuatan hukum tetap / inkracht) originating from district courts across five provinces in Java—West Java (Jawa Barat), Central Java (Jawa Tengah), East Java (Jawa Timur), DKI Jakarta, and DI Yogyakarta—spanning decision years 2005 through 2019.[1] Each XML file corresponds to a single court decision and contains structured header metadata alongside eleven segmented document sections annotated according to Supreme Court Decision Template Number 44/KMA/SK/III/2014: document opener (kepala putusan), defendant identity (identitas terdakwa), detention history (riwayat penahanan), indictment history (riwayat dakwaan), prosecution history (riwayat tuntutan), facts (fakta), legal facts (fakta hukum), legal considerations (pertimbangan hukum), verdict (amar putusan), verdict closing (penutup), and annexes.[1] All text parsing, rule-based feature extraction, quality-control auditing, and econometric modeling were executed locally in Python (analysis/indolaw_petty_theft/run_analysis.py and IndoLaw_low_value_theft_analysis.ipynb) on 22 October 2026 without transmitting any court decision text to external large language model APIs. Exported analytical datasets (analytic_cohort_deidentified.csv, strict_legal_relevance_manifest_deidentified.csv, and results.json) were restricted to cryptographic file hash identifiers (case_id), aggregate metadata, derived categorical indicators, and numerical variables, with zero extraction of personal names, residential addresses, or verbatim free-text quotations.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue X October 2026

Cohort Selection, Contextual Loss Extraction, and OCR Conflict Reconciliation

Cohort construction proceeded through a deterministic screening cascade across all N = 22,630 XML documents in the Indo-Law repository.[1,26] First, decisions were filtered using the structured metadata attribute sub_klasifikasi = pencurian (theft), isolating n = 2,398 theft decisions (10.60% of the corpus) and excluding 20,232 non-theft criminal decisions (89.40%). Second, within the 2,398 theft-labeled documents, monetary loss amounts denominated in Indonesian Rupiah (IDR) were extracted from contextual windows surrounding explicit loss lexemes (kerugian, rugi, senilai, seharga) within the indictment history (<riwayat_dakwaan>), trial facts (<fakta>), and legal facts (<fakta_hukum>) sections.[1,3,19] Because Indonesian court decisions frequently record monetary figures both as Arabic digits (for example, Rp 1.500.000) and as parenthetical Indonesian word-spelled numerals (for example, satu juta lima ratus ribu rupiah), and because PDF-to-text OCR conversion frequently introduces punctuation corruption or dropped trailing zeroes,[1,3] a dedicated bilingual lexical-numeric reconciliation engine audited all 8,828 digit–word pairs detected across the 2,398 theft files. Across these 8,828 compared pairs, 933 pairs exhibited any numerical difference between the digit string and the written numeral. When the ratio between the smaller and larger value lay within the tolerable minor rounding band of 0.80 to 1.25 (corresponding to a ≤ 20%–25% discrepancy, observed in 5 primary cohort documents), the pair was reconciled to the Indonesian word-spelled numeral. Conversely, whenever a digit string and its adjacent word-spelled numeral diverged beyond the 0.80–1.25 ratio band (> 20%–25% discrepancy), the pair was flagged as a material OCR conflict (identifying 734 material conflict pairs across 351 documents), and the document was excluded from the primary cohort to prevent high-value thefts with truncated digits from falsely entering the low-value sample. Furthermore, any decision containing explicit lexical qualifiers indicating losses above the threshold (lebih dari, di atas, atau melebihi; 411 matches across the theft sub-corpus) was excluded.

Two nested operational low-value theft cohorts were defined a priori using the IDR 2,500,000 nominal ceiling of Supreme Court Regulation (Perma) Number 2 of 2012 as an operational historical boundary.[4,11–14] The Primary Analytic Cohort (N = 148 decisions) required that: (1) at least one contextual monetary loss value was readable; (2) the maximum detected contextual loss value in the decision (loss_max_rupiah) did not exceed IDR 2,500,000; (3) the document contained zero material OCR conflicts; and (4) the document contained zero above-threshold exclusionary phrases. To evaluate sensitivity against multi-item or multi-count thefts in which individual stolen items each cost ≤ IDR 2,500,000 but their cumulative total exceeded IDR 2,500,000, a Conservative Sensitivity Cohort (n = 131 decisions; 88.51% of the primary cohort) was defined by requiring that the arithmetic sum of all unique monetary values detected across the loss context (loss_sum_unique_rupiah) also did not exceed IDR 2,500,000. Across both cohorts, the IDR 2,500,000 cutoff was treated strictly as an operational nominal loss proxy rather than a juridical classification of statutory minor theft under Article 364 KUHP.[4,15]

Operationalization of Outcome, Exposures, and Structural Case Covariates

All analytical variables were operationalized using deterministic rule-based extraction prior to statistical modeling.[1,20,26] The primary outcome variable was custodial sentence length measured in days (sentence_days), parsed directly from the structured Indo-Law metadata attribute lama_hukuman (which standardizes judicial prison sentences into days using 1 month = 30 days and 1 year = 360 or 365 days per repository convention).[1,27] For the primary linear regression model, sentence_days was transformed using the natural logarithm, ln(sentence_days), among decisions with positive custodial sentences (> 0 days; n = 147 of 148). Two binary secondary outcome indicators were also defined: (1) sentence length strictly exceeding 90 days (> 90 days), evaluated strictly as an empirical diagnostic benchmark against the 3-month maximum imprisonment ceiling of Article 364 KUHP minor theft rather than as a measure of statutory non-compliance;[4,11] and (2) sentence length strictly exceeding the complete-case cohort median (> 540 days).

The first primary exposure was recorded legal representation status (counsel_status), extracted from procedural history and header sections and classified into three mutually exclusive categories: assisted/represented (didampingi, when the decision explicitly recorded the presence or appointment of an advocate, legal aid counsel, or public defender), unassisted/unrepresented (tidak didampingi, when the decision explicitly stated that the defendant was not accompanied by legal counsel or waived representation), and unrecorded (tidak tercatat, when the XML text lacked explicit confirmation of either representation or non-representation).[7,28] The second primary

exposure was the defendant's socioeconomic proxy measured via occupational category (occupation_group), extracted from the defendant identity section (<identitas>, present as a segmented XML tag in 129/148 [87.16%] decisions and absent/unsegmented in 19/148 [12.84%]) using a lexical dictionary and classified into seven detailed strata: laborers and agriculture (buruh dan pertanian), business and private sector (usaha dan sektor swasta), unemployed (tidak bekerja), students or university students (pelajar atau mahasiswa), homemakers (rumah tangga), other formal sector (formal lainnya), and mixed or unrecorded occupation (campuran atau tidak tercatat).[1,8] For multivariable OLS modeling (n = 120), to avoid unstable parameter estimation in sparse cells (students n = 4, homemakers n = 2, other formal n = 1, unemployed n = 13 in N = 148), the five non-reference, non-business categories were consolidated a priori into 'Other or unrecorded occupation' (n = 54/120 [45.0%]) and compared against 'Business and private sector' (n = 36/120 [30.0%]) and the reference category 'Laborers and agriculture' (n = 30/120 [25.0%]).

Pre-specified structural case covariates comprised: (1) prosecutorial sentence demand in days (prosecution_demand_days), extracted from the prosecution history section (<riwayat_tuntutan>) and natural log-transformed as ln(prosecution_demand_days);[1,6] (2) statutory charge indicators for Article 362 (pasal_362), Article 363 (pasal_363), Article 364 (pasal_364), and Article 365 (pasal_365) KUHP, together with the composite aggravation or violence indicator (aggravated_or_violent), coded 1 if the indictment or verdict referenced Article 363, Article 365, or explicit lexical phrases denoting aggravated theft (pencurian dengan pemberatan) or violent theft (pencurian dengan kekerasan), and 0 otherwise;[15] (3) defendant multiplicity (multiple_defendants), coded 1 if two or more co-defendants (terdakwa jamak) were adjudicated and 0 for a single defendant; (4) centered decision calendar year (year_centered = year − 2016.0, centered at the convicted-cohort median year of 2016 so that the regression intercept represents a baseline case in 2016); and (5) province of the adjudicating district court (province: West Java [reference, n = 30 in complete cases], Central Java [n = 41], East Java [n = 31], DI Yogyakarta [n = 15], and DKI Jakarta [n = 3]).[1] Additionally, among the subset of decisions containing a segmented legal considerations section (<pertimbangan_hukum>; n = 80 of 148 [54.05%]), six binary judicial consideration indicators were extracted: four mitigating factors (admission of guilt or remorse [pengakuan atau penyesalan], polite courtroom demeanor [bersikap sopan], family dependents [tanggungan keluarga], and restitution or victim reconciliation [restitusi atau perdamaian]) and two aggravating factors (prior criminal conviction [pernah dihukum] and public unrest or community anxiety [meresahkan masyarakat]).[10,29–34]

Statistical Analysis Plan, Missing-Data Diagnostics, and Manual Validation Protocol

Continuous distributions were summarized non-parametrically using the median, interquartile range (IQR: 25th to 75th percentile), and full range (minimum to maximum), while categorical variables were reported as exact counts (n), percentages (%), and 95% Wilson score confidence intervals (CIs).[26,27] Unadjusted bivariate comparison of sentence length between the assisted (n = 19) and unassisted (n = 70) legal representation groups was performed using the two-sided Mann–Whitney U test.[27] Bivariate comparison of the binary proportion receiving sentences > 90 days across assisted versus unassisted status was evaluated using the two-sided Fisher's exact test, accompanied by Haldane–Anscombe continuity-corrected (+0.5 per cell) odds ratio estimation to document zero-cell boundary behavior when 100.0% (70/70) of unassisted defendants received > 90 days.

The primary multivariable model was an ordinary least squares (OLS) linear regression predicting natural log-transformed sentence length, ln(sentence_days), among complete-case decisions with positive sentences and extractable prosecutorial demands (n = 120; 81.08% of the primary cohort), estimated with MacKinnon–White HC3 heteroskedasticity-consistent standard errors.[28,35,36] Log-scale regression coefficients (β) and their 95% HC3 CIs for categorical, binary, and annual predictors were exponentiated into relative percentage differences in custodial sentence length, Δ% = (exp(β) − 1) × 100%; to ensure complete numerical transparency, both exact floating-point transformations [(exp(β_exact) − 1) × 100%] and 3-decimal rounded transformations [(exp(β_3dec) − 1) × 100%] were tabulated. For the continuous log-transformed prosecutorial demand variable, β represents the adjusted log–log association slope; the relative sentence change associated with a +10% increase in prosecutorial demand was computed both via exact power transformation [(1.10^β − 1) × 100% = +8.25%, or ~8.2%] and via first-order linear elasticity scaling (0.10 × β × 100% = +8.3%).

As a pre-specified secondary multivariable analysis, a binary logistic regression model with HC3 robust standard

errors was estimated on the complete-case cohort (n = 120) predicting custodial sentences strictly above the cohort median (> 540 days; McFadden pseudo-R^2 = 0.453). To prevent quasi-complete separation across sparse provincial (DKI Jakarta n = 3) and occupational cells in binary outcome modeling, the logistic specification adjusted for counsel status (assisted and unrecorded vs. unassisted reference), log prosecutorial demand, and centered calendar year. Finally, three quality-control and missingness audits were executed: (1) a deterministic simple random sample of n = 40 primary cohort decisions (27.03% of 148) drawn using fixed pseudo-random seed random_state = 20261022 to manually verify the sampled low-value loss phrase against the XML text; (2) non-parametric Mann–Whitney U missing-data diagnostics comparing sentence distributions between complete cases (n = 120) and excluded missing-demand/zero-sentence cases (n = 28), and between decisions with (n = 80) versus without (n = 68) segmented judicial considerations; and (3) a corpus-wide lexical search across all 22,630 Indo-Law XML files for explicit mentions of Article 364 KUHP (Pasal 364), minor theft (pencurian ringan), or Perma Number 2 of 2012 (Perma 2/2012).[6,26,28]

RESULTS

Corpus Screening Cascade, Quality Control Audit, and Primary Cohort Structure

Screening the entire Indo-Law repository of N = 22,630 structured XML criminal court decisions (commit 67340338adbc63021630947d905627eef0d7e43f; decision years 2005–2019 across five provinces in Java) identified n = 2,398 decisions (10.60% of the corpus; 95% Wilson CI: 10.20%–11.01%) carrying the sub-classification label pencurian (theft), while 20,232 decisions (89.40%) pertained to other criminal offenses, as detailed in Table 1. Among the 2,398 theft-labeled decisions, contextual monetary loss figures were parsed from the indictment history, trial facts, or legal facts sections in n = 2,138 decisions (89.16% of theft decisions; 9.45% of the entire corpus), whereas 260 theft decisions (10.84%) lacked extractable monetary loss values. Across the 2,398 theft files, the bilingual reconciliation engine compared 8,828 digit–word numeral pairs, identifying 933 pairs with any numerical difference: minor differences within the tolerable 0.80–1.25 ratio band (occurring in 5 primary cohort documents) were resolved to the written numeral, whereas 734 material OCR conflict pairs (< 0.80 or > 1.25 ratio) distributed across 351 documents (16.42% of the 2,138 readable-loss theft decisions) and 411 above-threshold qualifier matches (lebih dari, di atas, melebihi) were excluded. Applying the primary operational inclusion criteria—all contextual monetary losses ≤ IDR 2,500,000, zero material OCR conflicts, and zero exclusionary above-threshold phrases—yielded a Primary Analytic Cohort of N = 148 decisions (6.92% of readable-loss theft decisions; 6.17% of all theft decisions; 0.65% of the 22,630-document repository), with zero duplicate case hash identifiers (0.0%). Applying the stricter conservative sensitivity definition—requiring that the cumulative sum of all unique monetary values detected in the loss context also not exceed IDR 2,500,000—retained n = 131 decisions (88.51% of the primary cohort). Deterministic random manual auditing of n = 40 primary cohort decisions (27.03% of 148; drawn using random_state = 20261022) confirmed that 40 of 40 sampled decisions (100.0%; 95% CI: 91.2%–100.0%) contained an explicit low-value monetary loss phrase ≤ IDR 2,500,000 in the sampled context (with the explicit methodological caveat that this audit verifies the sampled low-value phrase rather than proving exhaustive absence of uncaptured amounts elsewhere in unstructured narrative sections), as detailed in Table 1.

Table 1. Corpus Screening Cascade, Operational Variable Definitions, and Quality Control Audit Across the Indo-Law XML Repository (N = 22,630 Decisions; 2005–2019).

Screening Stage / Analytical Domain	Operational Rule & XML Source Section	Document / Pair Volume (n)	Proportion of Parent Stratum (%)	Methodological Rationale & Specificity Guardrail
Total Indo-Law XML Repository	First-instance general criminal decisions with permanent legal force across West Java, Central Java, East Java, DKI Jakarta, and DI Yogyakarta (2005–2019)	22,630	100.0% of corpus	Provides structured 11-section XML annotations enabling local reproducible parsing without external model transmission

Theft Sub-Classification Filter	Metadata attribute sub_klasifikasi = pencurian in verdict header	2,398	10.60% of corpus (2,398 / 22,630)	Isolates property theft prosecutions from non-theft criminal offenses (n = 20,232; 89.40%)
Readable Monetary Loss Extraction	Contextual regex and lexical extraction around loss terms (kerugian) in indictment history (<riwayat_dakwaan>), facts (<fakta>), and legal facts (<fakta_hukum>)	2,138	89.16% of theft (2,138 / 2,398; 9.45% of corpus)	Excludes 260 theft decisions (10.84%) lacking extractable monetary loss figures in narrative sections
Digit–Word Numeral OCR Reconciliation & Conflict Audit	8,828 digit–word pairs audited across 2,398 theft files (933 any-mismatch pairs). Minor differences within 0.80–1.25 ratio (5 primary cohort files) resolved to written numeral; material divergence (<0.80 or >1.25 ratio) or >threshold qualifiers (411 matches: lebih dari, di atas, melebihi) excluded	734 material conflict pairs in 351 documents	16.42% of readable-loss theft (351 / 2,138; 14.64% of theft)	Eliminates OCR truncation artifacts where multi-million Rupiah losses lose trailing zeroes during PDF-to-text conversion
Primary Analytic Cohort	All detected contextual loss values ≤ IDR 2,500,000, zero material OCR conflicts, and zero phrases indicating losses above threshold	148	6.92% of readable loss (6.17% of theft; 0.65% of corpus)	Operational historical proxy for nominally low-value theft aligned with the IDR 2,500,000 threshold of Perma No. 2/2012
Conservative Sensitivity Cohort	Sum of all unique detected monetary values across the decision ≤ IDR 2,500,000 with zero material OCR conflicts	131	88.51% of primary cohort (131 / 148; 6.13% of readable loss)	Tests robustness against multi-item or multi-victim theft where cumulative loss could exceed the single-item maximum
Deterministic Random Manual Audit & Duplicate Check	Direct human verification of a simple random sample of 40 primary cohort decisions (random_state = 20261022) and corpus-wide unique hash ID check	40 / 40 confirmed; 0 duplicates	27.03% of primary cohort audited; 100.0% phrase precision (95% CI: 91.2%–100.0%)	Confirms sampled explicit low-value loss phrase (≤ IDR 2,500,000) in all 40 audited decisions (does not prove absence of uncaptured values in unstructured sections)
Outcome, Exposure & Covariate Domains	Outcome: sentence_days (lama_hukuman; ln-transformed for >0d). Exposures: counsel_status (3 strata) and occupation_group (7 lexical strata; collapsed to 3 in OLS; <identitas> present in 129/148	148 primary (120 complete-case; 80 considerations)	81.08% complete-case (120 / 148); 54.05% considerations (80 / 148)	Pre-specified covariate adjustment isolates associations of counsel recording and occupation from

	[87.16%]). Covariates: ln(prosecution_demand_days), aggravated_or_violent (Art. 363/365 or phrase), multiple_defendants, year_centered (year − 2016.0), and province			prosecutorial demand and offense severity

Abbreviations: CI, Wilson score confidence interval; IDR, Indonesian Rupiah; OCR, optical character recognition; OLS, ordinary least squares; Perma, Supreme Court Regulation (Peraturan Mahkamah Agung); XML, Extensible Markup Language.

The five-stage computational screening architecture and the quantitative document attrition cascade across the repository are illustrated in Figure 1 (Panel a and Panel b). As shown in Figure 1a, auditing all 8,828 digit–word pairs prevented OCR digit truncation errors from contaminating the low-value cohort, while the logarithmic attrition profile in Figure 1b demonstrates the progressive narrowing from 22,630 total decisions to the 148 primary cohort cases, 131 conservative sensitivity cases, 120 complete-case multivariable cases, 80 structured consideration cases, 40 manually audited cases, and 5 corpus-wide explicit Article 364 / Perma Number 2 of 2012 cases.

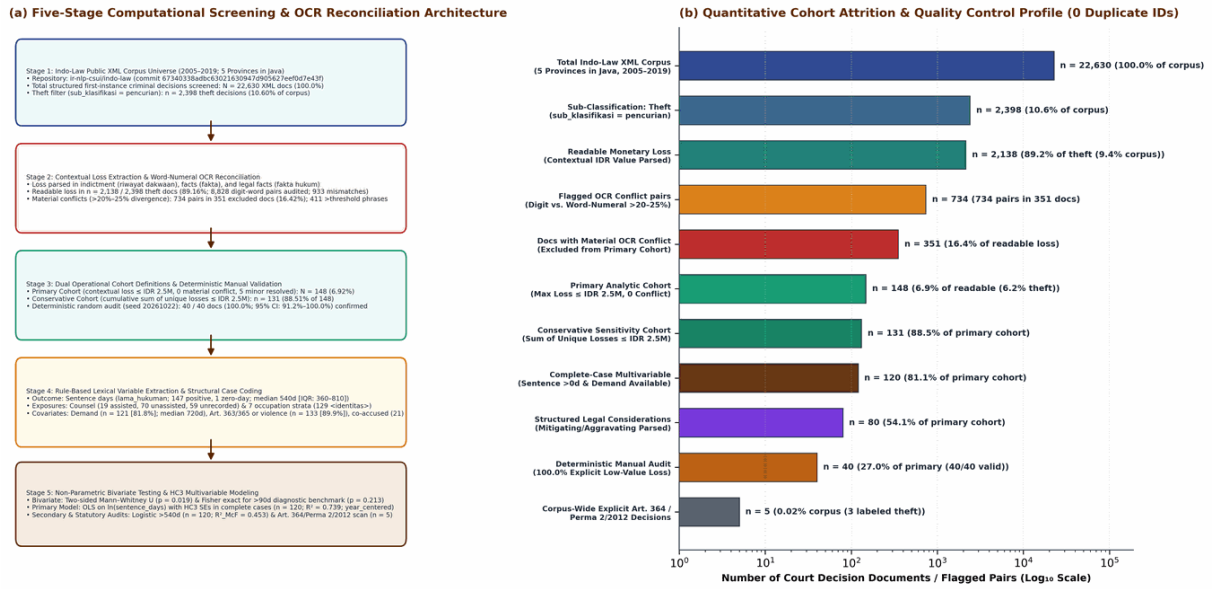

Figure 1. Five-stage computational screening and rule-based lexical-numeric OCR reconciliation architecture across the Indo-Law XML repository (8,828 digit–word pairs audited across 2,398 theft files; 734 material conflict pairs in 351 excluded files; 411 above-threshold qualifiers) (Panel a), and quantitative cohort attrition cascade from N = 22,630 total decisions to the primary low-value theft cohort (n = 148), conservative sensitivity cohort (n = 131), complete-case multivariable cohort (n = 120), structured legal considerations subset (n = 80), deterministic manual validation sample (n = 40; random_state = 20261022), and corpus-wide explicit Article 364 / Perma No. 2/2012 hits (n = 5) (Panel b).

Across the N = 148 decisions in the primary low-value theft cohort (distributed across Central Java [n = 48, 32.4%], West Java [n = 39, 26.4%], East Java [n = 39, 26.4%], DI Yogyakarta [n = 19, 12.8%], and DKI Jakarta [n = 3, 2.0%]), custodial sentence length (sentence_days) was right-skewed, with a cohort median of 540 days (18 months), an interquartile range (IQR) of 360 to 810 days (12 to 27 months; IQR width = 450 days), and a full range spanning 0 to 4,320 days (0 to 12 years), as detailed in Table 2. Exactly 147 decisions (99.32%, reported as 99.3%; 95% CI: 96.3%–99.9%) imposed a positive custodial prison sentence (> 0 days), while 1 decision (0.68%, reported as 0.7%) recorded a non-custodial disposition of 0 days (lepas). Evaluated against the 90-day (3-month) diagnostic benchmark of Article 364 KUHP, 146 of the 148 decisions (98.65%, reported as 98.6%; 95% CI: 95.2%–99.6%) imposed custodial sentences strictly exceeding 90 days, leaving only 2 decisions

(1.35%, reported as 1.4%) at or below 90 days (one at 0 days and one at exactly 90 days). Statutory charge breakdown explained why these cases were adjudicated as ordinary or aggravated felonies rather than summary minor offenses: 0 of 148 primary cohort decisions (0.0%) cited Article 364 KUHP, whereas 79 decisions (53.38%, reported as 53.4%) explicitly cited Article 363 KUHP (aggravated theft), 66 decisions (44.59%, reported as 44.6%) cited Article 365 KUHP (theft with violence), 35 decisions (23.65%, reported as 23.6%) cited Article 362 KUHP (ordinary theft), and 133 of the 148 decisions (89.86%, reported as 89.9%; 95% CI: 83.9%–93.8%) satisfied the composite aggravation or violence indicator (aggravated_or_violent). Multiple co-defendants (terdakwa jamak) were prosecuted in 21 decisions (14.19%, reported as 14.2%; 95% CI: 9.5%–20.7%), versus a single defendant in 127 decisions (85.81%, reported as 85.8%).

Prosecutorial sentence demand in days (prosecution_demand_days) was extractable in 121 decisions (81.76%, reported as 81.8%; 95% CI: 74.8%–87.2%), exhibiting a median demand of 720 days (24 months; IQR: 360–720 days in n = 121; IQR: 360–810 days in the n = 120 complete-case cohort; full range: 90–3,240 days), and was missing in 27 decisions (18.24%, reported as 18.2%), as detailed in Table 2. Combining positive sentence length (> 0 days) with available prosecutorial demand yielded n = 120 complete-case decisions (81.08%, reported as 81.1% of 148; median sentence = 540 days, IQR: 322.5–720 days) for multivariable regression modeling. Quantitative missing-data comparisons confirmed no statistically significant difference in custodial sentence length between the n = 120 complete-case decisions (median 540 days, IQR: 322.5–720; aggravated/violent 90.8%) and the n = 28 excluded decisions (median 585 days, IQR: 450–900; aggravated/violent 85.7%; two-sided Mann–Whitney U test p = 0.249), nor between the n = 80 decisions with a segmented <pertimbangan_hukum> section (median 540 days, IQR: 360–817.5) and the n = 68 decisions without one (median 615 days, IQR: 360–810; Mann–Whitney U test p = 0.890), as summarized in Table 2.

Table 2. Baseline Characteristics, Statutory Offense Composition, Prosecutorial Demand Alignment, and Missing-Data Profile of the Primary Low-Value Theft Cohort (N = 148 Decisions; 2005–2019).

Cohort Characteristic / Parameter	**Operational Measurement & Coding Scale**	**Observed Summary Value / Count (n)**	**Percentage or Distributional Spread**	**95% Wilson CI or Comparative Metric**
Primary Analytic Sample Size	Eligible low-value theft decisions (explicit loss ≤ IDR 2,500,000) across five provinces in Java (2005–2019)	N = 148	100.0% of primary cohort (West Java 39, Central Java 48, East Java 39, DI Yogyakarta 19, DKI Jakarta 3)	Zero duplicate case hash identifiers (0.0%); <identitas> present in 129/148 (87.2%)
Custodial Sentence Length (sentence_days)	Duration of imprisonment in days extracted from structured verdict metadata (lama_hukuman)	Median = 540 days	IQR: 360–810 days (IQR width: 450 days)	Full range (min–max): 0–4,320 days (complete-case n = 120 median = 540 days, IQR: 322.5–720)
Positive vs. Zero-Day Custodial Sentence	Positive custodial sentence (> 0 days) versus non-custodial / zero-day verdict (= 0 days; lepas)	147 positive; 1 zero-day	99.3% positive; 0.7% zero-day (Total = 100.0%)	Positive sentence 95% CI: 96.3%–99.9%

Sentence Exceeding 90-Day Diagnostic Benchmark	Custodial sentence strictly exceeding the 90-day (3-month) statutory ceiling of Article 364 KUHP minor theft (diagnostic benchmark; 0/148 cited Art. 364)	146 (> 90 days); 2 (≤ 90 days)	98.6% (> 90 days); 1.4% (≤ 90 days) (Total = 100.0%)	> 90 days 95% CI: 95.2%–99.6% (≤ 90d: 1 case = 0d [lepas], 1 case = 90d)
Statutory Charge Breakdown & Composite Aggravation (aggravated_or_violent)	Explicit citation of Article 362 (ordinary theft), Article 363 (aggravated theft), Article 364 (minor theft), Article 365 (violent theft), or composite aggravation/violence indicator	133 composite present; 15 absent	Art. 363: 79 (53.4%); Art. 365: 66 (44.6%); Art. 362: 35 (23.6%); Art. 364: 0 (0.0%); Composite: 89.9%	Composite aggravated/violent 95% CI: 83.9%–93.8% (non-mutually exclusive statutory citations)
Defendant Multiplicity (multiple_defendants)	Multiple co-defendants (terdakwa jamak) versus single defendant prosecuted in decision	21 multiple; 127 single	14.2% multiple; 85.8% single (Total = 100.0%)	Multiple defendants 95% CI: 9.5%–20.7%
Prosecutorial Sentence Demand & Sentence-to-Demand Ratio	Extractable prison demand in days (prosecution_demand_days) from <riwayat_tuntutan> and empirical alignment with imposed sentence	121 available (Med = 720d); 27 missing	81.8% available (IQR: 360–720d in n=121; 360–810d in n=120; range: 90–3,240d); 18.2% missing	Available 95% CI: 74.8%–87.2%; Median sentence/demand ratio = 0.866 (86.6%, IQR: 69.2%–100.0%); 61.7% (74/120) < demand, 15.0% (18/120) = demand, 23.3% (28/120) > demand
Complete-Case vs. Excluded Missing-Demand Diagnostic	Comparison of complete-case multivariable cohort (n = 120; sentence > 0d & demand available) vs. excluded decisions (n = 28; 27 missing demand, 1 zero-day sentence)	n = 120 complete; n = 28 excluded	81.1% complete (Med sentence = 540d, IQR: 322.5–720; aggravated 90.8%) vs. 18.9% excluded (Med = 585d, IQR: 450–900; aggravated 85.7%)	Complete-case 95% CI: 74.0%–86.6%; Complete vs. excluded sentence Mann–Whitney U p = 0.249 (non-significant)
Structured Judicial Considerations Availability & Missingness Audit	Presence of segmented <pertimbangan_hukum> section (n = 80) vs. unsegmented/missing section (n = 68)	80 available; 68 unstructured	54.1% available (Med sentence = 540d, IQR: 360–817.5) vs. 45.9% unstructured (Med = 615d, IQR: 360–810)	Available 95% CI: 46.0%–61.9%; Available vs. unstructured sentence Mann–Whitney U p =

				0.890 (non-significant)

Abbreviations: CI, Wilson score confidence interval; IDR, Indonesian Rupiah; IQR, interquartile range (25th to 75th percentile); KUHP, Indonesian Criminal Code (Kitab Undang-Undang Hukum Pidana).

Recorded Legal Representation and Bivariate Custodial Sentence Distributions

Within the primary cohort of N = 148 low-value theft decisions, procedural recording of legal counsel (counsel_status) classified 19 decisions (12.84%, reported as 12.8%) as assisted/represented (didampingi), 70 decisions (47.30%, reported as 47.3%) as explicitly unassisted/unrepresented (tidak didampingi), and 59 decisions (39.86%, reported as 39.9%) as unrecorded (tidak tercatat), summing to 100.0%, as detailed in Table 3. In unadjusted non-parametric comparisons, recorded legal representation exhibited a pronounced crude gradient in custodial sentence duration: defendants with recorded legal counsel received a median sentence of 300 days (10 months; IQR: 165–585 days; IQR width = 420 days), compared with a median of 600 days (20 months; IQR: 375–787 days; IQR width = 412 days) among explicitly unassisted defendants and a median of 630 days (21 months; IQR: 450–900 days; IQR width = 450 days) among defendants with unrecorded counsel status. The crude 300-day (50.0%) median difference between the assisted group (n = 19; median 300 days) and the explicitly unassisted group (n = 70; median 600 days) was statistically significant on a two-sided Mann–Whitney U test (p = 0.019). Critically, however, cross-tabulating prosecutorial sentence demands by counsel status within the complete-case cohort (n = 120) revealed that assisted defendants (n = 16; median sentence 285 days, IQR: 146.2–540) already faced a median prosecutorial demand of 360 days (IQR: 232.5–360 days), which was exactly half the median prosecutorial demand of 720 days faced by both unassisted defendants (n = 57; demand IQR: 360–720 days; median sentence 600 days) and unrecorded-counsel defendants (n = 47; demand IQR: 360–1,080 days; median sentence 600 days), as detailed in Table 3.

Table 3. Bivariate Distribution of Sentence Length in Days, Diagnostic 90-Day Benchmark Exceedance, and Complete-Case Prosecutorial Demands Stratified by Recorded Legal Representation Status (N = 148 Primary Cohort; n = 120 Complete Cases).

Recorded Legal Counsel Status (counsel_status)	**Cohort Count & Share (N = 148 / n = 120)**	**Median Sentence (Days, N = 148 / n = 120)**	**Sentence IQR (Days, N = 148) & Demand Median (IQR, n = 120)**	**Sentence > 90-Day Benchmark (n/N, % [95% Wilson CI])**	**Bivariate Hypothesis Tests & Effect Size vs. Unassisted Reference**
Assisted / Represented (Didampingi)	19 / 148 (12.8%) [CC: 16 / 120 (13.3%)]	300d (N = 148) [CC Med: 285d]	Sentence IQR: 165–585d (CC: 146.2–540d) Demand Med (n=16): 360d (IQR: 232.5–360d)	18 / 19 (94.7% [75.4%–99.1%]); 1 / 19 (5.3%) ≤ 90d	Crude median difference = −300 days (−50.0%); two-sided Mann–Whitney U test p = 0.019*; Fisher's exact test for >90d p = 0.213 (Haldane–Anscombe OR = 0.09 [95% CI: 0.003–2.24])
Unassisted / Unrepresented (Tidak didampingi) [Reference]	70 / 148 (47.3%) [CC: 57 / 120 (47.5%)]	600d (N = 148) [CC Med: 600d]	Sentence IQR: 375–787d (CC: 360–720d) Demand Med (n=57): 720d	70 / 70 (100.0% [94.8%–100.0%]); 0 / 70 (0.0%) ≤ 90d	Reference category for bivariate Mann–Whitney U test, Fisher's exact test, and multivariable

			(IQR: 360–720d)		regression models; crude unconditional OR for >90d undefined due to zero ≤90d cell
Unrecorded Status (Tidak tercatat)	59 / 148 (39.9%) [CC: 47 / 120 (39.2%)]	630d (N = 148) [CC Med: 600d]	Sentence IQR: 450–900d (CC: 405–765d) Demand Med (n=47): 720d (IQR: 360–1,080d)	58 / 59 (98.3% [91.0%–99.7%]); 1 / 59 (1.7%) ≤ 90d	Crude median difference vs. unassisted = +30 days (+5.0%); reflects administrative omission or unparsed procedural text rather than confirmed absence of counsel
Total Cohort Summary	148 (100.0%) [CC: 120 (100.0%)]	540d (N = 148) [CC Med: 540d]	Sentence IQR: 360–810d (CC: 322.5–720d) Demand Med (n=120): 720d (IQR: 360–810d)	146 / 148 (98.6% [95.2%–99.6%]); 2 / 148 (1.4%) ≤ 90d	Assisted cases entered sentencing with half the median prosecutorial demand of unassisted cases (360 vs. 720 days); crude counsel difference attenuates to −8.9% (p = 0.477) in adjusted OLS

**Statistically significant at α = 0.05 in unadjusted two-sided Mann–Whitney U comparison between assisted (n = 19) and unassisted (n = 70) decisions. Note: The 90-day (3-month) threshold represents the statutory maximum of Article 364 KUHP minor theft and is evaluated strictly as an empirical diagnostic benchmark (0/148 primary cohort decisions cited Article 364). Abbreviations: CC, complete-case multivariable cohort (n = 120); CI, confidence interval; IQR, interquartile range; OR, odds ratio.*

The non-parametric custodial sentence distributions and the diagnostic 90-day benchmark profile across legal representation strata are illustrated in Figure 2 (Panel a and Panel b). As depicted in Figure 2a, although the interquartile box for assisted defendants shifts downward toward shorter prison durations (median 300 days vs. 600 days in unassisted decisions), nearly the entire distribution across all three groups lies well above the 90-day diagnostic benchmark of Article 364 KUHP. When evaluating the binary proportion of decisions exceeding 90 days, as shown in Figure 2b, the difference across representation strata was small and imprecise: sentences > 90 days were imposed in 18 of 19 assisted decisions (94.74%, reported as 94.7%; 95% CI: 75.4%–99.1%; 1 decision [5.3%] ≤ 90 days), 70 of 70 explicitly unassisted decisions (100.0%; 95% CI: 94.8%–100.0%; 0 decisions ≤ 90 days), and 58 of 59 unrecorded-status decisions (98.31%, reported as 98.3%; 95% CI: 91.0%–99.7%; 1 decision [1.7%] ≤ 90 days). Two-sided Fisher's exact comparison of sentences > 90 days between assisted (18/19) and unassisted (70/70) defendants yielded p = 0.213; because 100.0% of the 70 explicitly unassisted decisions exceeded 90 days (creating a zero cell for ≤ 90 days), the unconditional crude odds ratio was mathematically undefined (Haldane–Anscombe continuity-corrected OR = 0.09, 95% CI: 0.003–2.24), as illustrated in Figure 2.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue X October 2026

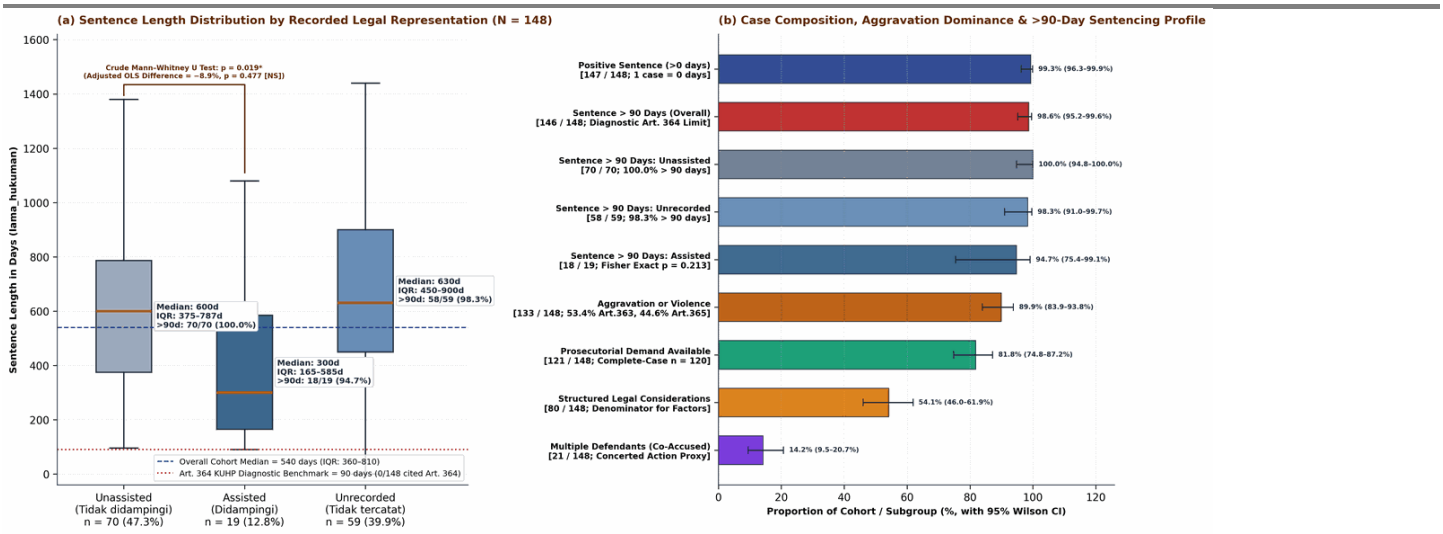

Figure 2. Non-parametric custodial sentence length distributions (median, interquartile range box, and whiskers in days) across recorded legal representation strata (n = 70 unassisted, n = 19 assisted, and n = 59 unrecorded; crude Mann–Whitney U test p = 0.019 for assisted vs. unassisted) relative to the cohort median (540 days) and the 90-day Article 364 KUHP diagnostic benchmark (0/148 cited Article 364) (Panel a), and structural case composition and >90-day profile with 95% Wilson confidence intervals across the primary cohort (N = 148) (Panel b).

Lexical Defendant Occupation Proxies and Judicial Considerations Profile

Lexical classification of defendant occupation (occupation_group) across all N = 148 primary cohort decisions revealed substantial missingness alongside sparse cell counts in minor demographic categories, as detailed in Table 4 (Panel A). The largest category was mixed or unrecorded occupation (campuran atau tidak tercatat), comprising 45 decisions (30.41%, reported as 30.4%, including 19 decisions lacking a segmented <identitas> tag), with a median sentence of 540 days (IQR: 360–720 days). Business and private-sector workers (usaha dan sektor swasta) formed the second largest group with 44 decisions (29.73%, reported as 29.7%), also exhibiting a median sentence of 540 days (IQR: 292–720 days). Manual laborers and agricultural workers (buruh dan pertanian) accounted for 39 decisions (26.35%, reported as 26.4%) and experienced a higher crude median sentence of 630 days (IQR: 450–900 days). Explicitly unemployed defendants (tidak bekerja) accounted for 13 decisions (8.78%, reported as 8.8%) and shared the same crude median sentence of 630 days (IQR: 450–900 days). The remaining three occupational strata contained very small cell sizes that produced unstable descriptive intervals: students or university students (pelajar atau mahasiswa; n = 4 [2.70%, reported as 2.7%], median = 255 days, IQR: 142–543 days), homemakers (rumah tangga; n = 2 [1.35%, reported as 1.4%], median = 285 days, IQR: 202–367 days), and other formal-sector employees (formal lainnya; n = 1 [0.68%, reported as 0.7%], median = 660 days, IQR: 660–660 days). In the complete-case multivariable cohort (n = 120), the three collapsed occupational categories comprised laborers and agriculture (n = 30 [25.0%]), business and private sector (n = 36 [30.0%]), and other or unrecorded occupation (n = 54 [45.0%]).

A segmented judicial considerations section (<pertimbangan_hukum>) was available and parseable in 80 of the 148 primary cohort decisions (54.05%, reported as 54.1%; missing or unsegmented in 68 decisions [45.9%]), as detailed in Table 4 (Panel B). Using n = 80 as the explicit denominator, the most prevalent mitigating consideration (hal meringankan) was the defendant's admission of guilt or expression of remorse (pengakuan atau penyesalan), recorded in 76 of 80 decisions (95.0%; 95% CI: 87.8%–98.0%; corresponding to 51.4% of all 148 cohort decisions), followed by polite courtroom demeanor (bersikap sopan) in 51 of 80 decisions (63.75%, reported as 63.7%; 95% CI: 52.8%–73.4%; 34.5% of total cohort). Socio-familial and restorative mitigating factors were recorded far less frequently: family dependents (tanggungan keluarga) appeared in 11 of 80 decisions (13.75%, reported as 13.8%; 95% CI: 7.9%–23.0%; 7.4% of total cohort), and financial restitution or victim reconciliation (restitusi atau perdamaian) was documented in only 7 of 80 decisions (8.75%, reported as 8.8%; 95% CI: 4.3%–17.0%; 4.7% of total cohort). Conversely, aggravating judicial considerations (hal memberatkan) were highly prevalent: causing public unrest or community anxiety (meresahkan masyarakat) was

cited in 65 of 80 decisions (81.25%, reported as 81.2%; 95% CI: 71.3%–88.3%; 43.9% of total cohort), and prior criminal conviction or recidivism (pernah dihukum) was cited in 57 of 80 decisions (71.25%, reported as 71.2%; 95% CI: 60.5%–80.0%; 38.5% of total cohort), as detailed in Table 4.

Table 4. Sentence Length Distribution Across Lexical Defendant Occupation Groups (N = 148 Primary Cohort and n = 120 Collapsed Regression Categories) and Prevalence of Judicial Aggravating and Mitigating Considerations (n = 80).

Analytical Panel / Variable Category	**Subgroup / Regression Mapping / Consideration Role**	**Observed Count (n)**	**Stratum Percentage (%)**	**Median Sentence (Days) or 95% Wilson CI**	**Interquartile Range (IQR, Days) or Cohort-Wide Share (% of N = 148)**
Panel A: Defendant Occupation Group (N = 148) — Mixed or unrecorded (Campuran atau tidak tercatat)	Unclassified / missing proxy → Collapsed into 'Other or unrecorded' in OLS (n = 54/120 [45.0%] total collapsed)	45	30.4% of 148	Median: 540 days	IQR: 360–720 days (includes 19/148 decisions without segmented <identitas>)
Panel A: Business and private sector (Usaha dan sektor swasta)	Self-employed / private-sector → Modeled as 'Business and private sector' in OLS (n = 36/120 [30.0%])	44	29.7% of 148	Median: 540 days	IQR: 292–720 days (modeled as distinct category vs. laborers & agriculture)
Panel A: Laborers and agriculture (Buruh dan pertanian)	Manual labor / agriculture → Reference category in multivariable OLS (n = 30/120 [25.0%])	39	26.4% of 148	Median: 630 days	IQR: 450–900 days (reference category in multivariable OLS)
Panel A: Unemployed (Tidak bekerja)	Explicitly recorded joblessness → Collapsed into 'Other or unrecorded' in OLS	13	8.8% of 148	Median: 630 days	IQR: 450–900 days (grouped into 'Other or unrecorded' to prevent sparse-cell instability)
Panel A: Students or university students (Pelajar atau mahasiswa)	Secondary or tertiary student status (sparse cell) → Collapsed into 'Other or unrecorded' in OLS	4	2.7% of 148	Median: 255 days	IQR: 142–543 days (small cell n = 4; unstable descriptive point estimate)
Panel A: Homemakers (Rumah tangga)	Domestic / household care status (sparse cell) → Collapsed into 'Other or unrecorded' in OLS	2	1.4% of 148	Median: 285 days	IQR: 202–367 days (small cell n = 2; unstable descriptive point estimate)

Panel A: Other formal sector (Formal lainnya)	Formal salaried occupation (sparse cell) → Collapsed into 'Other or unrecorded' in OLS	1	0.7% of 148	Median: 660 days	IQR: 660–660 days (single observation n = 1; Total Panel A = 148 [100.0%])
Panel B: Judicial Considerations (n = 80 available) — Admission of guilt or remorse (Pengakuan atau penyesalan)	Mitigating consideration (Hal meringankan)	76 / 80	95.0% of 80	95% CI: 87.8%–98.0%	51.4% of total cohort (76 / 148)
Panel B: Polite courtroom demeanor (Bersikap sopan)	Mitigating consideration (Hal meringankan)	51 / 80	63.7% of 80	95% CI: 52.8%–73.4%	34.5% of total cohort (51 / 148)
Panel B: Family dependents (Tanggungan keluarga)	Mitigating consideration (Hal meringankan)	11 / 80	13.8% of 80	95% CI: 7.9%–23.0%	7.4% of total cohort (11 / 148)
Panel B: Restitution or victim reconciliation (Restitusi atau perdamaian)	Mitigating consideration (Hal meringankan)	7 / 80	8.8% of 80	95% CI: 4.3%–17.0%	4.7% of total cohort (7 / 148)
Panel B: Public unrest / community anxiety (Meresahkan masyarakat)	Aggravating consideration (Hal memberatkan)	65 / 80	81.2% of 80	95% CI: 71.3%–88.3%	43.9% of total cohort (65 / 148)
Panel B: Prior criminal conviction / recidivism (Pernah dihukum)	Aggravating consideration (Hal memberatkan)	57 / 80	71.2% of 80	95% CI: 60.5%–80.0%	38.5% of total cohort (57 / 148)

Note: In Panel A, exact counts sum to 148 (30.41% + 29.73% + 26.35% + 8.78% + 2.70% + 1.35% + 0.68% = 100.00%; one-decimal rounded values sum to 100.1%). For multivariable OLS modeling (n = 120), the 5 non-reference/non-business categories (Mixed/unrecorded, Unemployed, Students, Homemakers, and Other formal sector) were collapsed a priori into 'Other or unrecorded occupation' (n = 54/120 [45.0%]) versus 'Business and private sector' (n = 36/120 [30.0%]) and 'Laborers and agriculture' [Reference, n = 30/120 (25.0%)]. In Panel B, percentages use the 80 decisions with a structured <pertimbangan_hukum> section as the denominator (68/148 [45.9%] lacked segmented considerations; median sentence 540d vs. 615d, Mann–Whitney p = 0.890). Abbreviations: CI, Wilson score confidence interval; IQR, interquartile range; OLS, ordinary least squares.

The comparative median sentence intervals across the seven lexical occupation groups and the prevalence hierarchy of mitigating versus aggravating judicial considerations are illustrated in Figure 3 (Panel a and Panel b). As shown in Figure 3a, the interquartile ranges of laborers/agricultural workers (450–900 days), unemployed defendants (450–900 days), business/private-sector workers (292–720 days), and mixed/unrecorded occupations

(360–720 days) overlap substantially around the 540-day cohort median, whereas Figure 3b highlights the contrast between near-universal courtroom remorse (95.0%) and rare victim restitution or reconciliation (8.8%) alongside frequent recidivism (71.2%) and community unrest citations (81.2%).

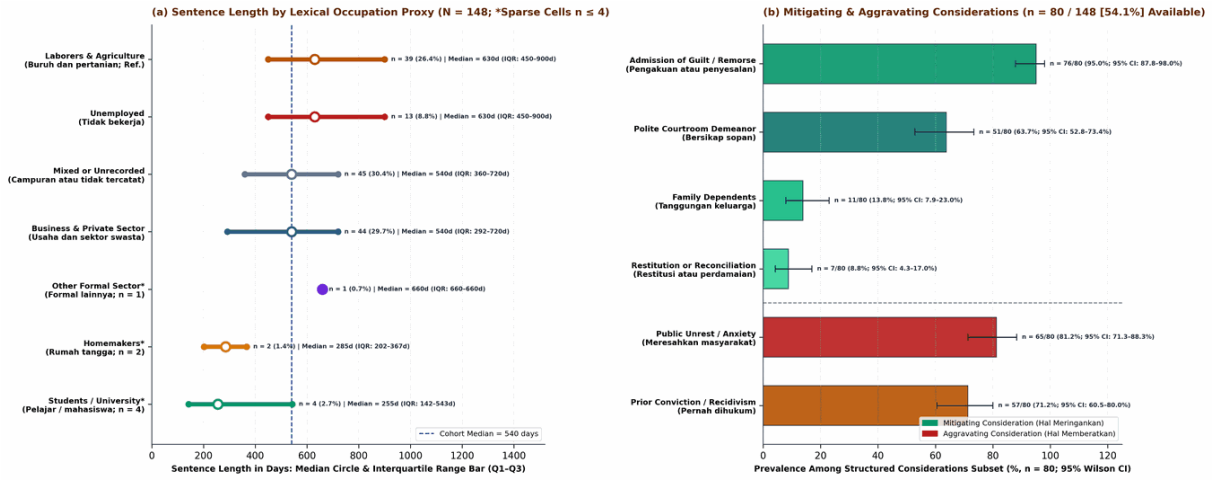

Figure 3. Custodial sentence length medians and interquartile ranges across the seven lexical defendant occupation groups in the primary cohort (N = 148; highlighting sparse cells with n ≤ 4 that are collapsed into 'Other or unrecorded' in multivariable modeling) (Panel a), and prevalence with 95% Wilson confidence intervals of mitigating and aggravating judicial considerations among the n = 80 decisions (54.1%) containing a structured <pertimbangan_hukum> section (Panel b).

Multivariable OLS Regression with HC3 Robust Standard Errors and Secondary Logistic Modeling

Estimating the pre-specified multivariable OLS regression model on natural log-transformed custodial sentence days with MacKinnon–White HC3 heteroskedasticity-consistent standard errors across the n = 120 complete-case decisions accounted for 73.9% of the total variance in log sentence duration (R^2 = 0.739; intercept at reference covariates in centered year 2016: β = 0.841 [0.8412], 95% CI: −0.010 to 1.693, p = 0.053), as detailed in Table 5. Crucially, once prosecutorial sentence demand, aggravation or violence, defendant multiplicity, centered decision year, and province were controlled, the crude sentencing difference observed for recorded legal representation vanished. Relative to explicitly unassisted defendants (tidak didampingi; n = 57), defendants with recorded legal representation (didampingi; n = 16) exhibited an adjusted log-scale coefficient of β = −0.093 (−0.0935 exact; 95% CI: −0.350 to 0.163; p = 0.477), corresponding to a statistically non-significant relative sentence difference of −8.9% (−8.92% exact; 95% CI: −29.5% to +17.8%). Similarly, defendants with unrecorded counsel status (tidak tercatat; n = 47) showed no independent difference relative to unassisted defendants, with β = 0.057 (0.0569 exact; 95% CI: −0.119 to 0.233; p = 0.527), corresponding to an adjusted sentence difference of +5.9% (+5.85% exact; 95% CI: −11.2% to +26.2%).

Collapsed occupational socioeconomic proxy categories likewise demonstrated no independent association with custodial sentence duration in the multivariable model, as detailed in Table 5. Compared against the reference group of manual laborers and agricultural workers (buruh dan pertanian; n = 30), defendants classified under other or unrecorded occupations (n = 54) exhibited β = 0.032 (0.0317 exact; 95% CI: −0.154 to 0.218; p = 0.738), equivalent to a +3.2% sentence difference (+3.22% from exact float; +3.3% from 3-decimal β = 0.032; 95% CI: −14.3% to +24.3%), whereas defendants in business and the private sector (usaha dan swasta; n = 36) exhibited β = 0.096 (0.0957 exact; 95% CI: −0.135 to 0.327; p = 0.417), equivalent to a +10.0% sentence difference (+10.05% from exact float; +10.1% from 3-decimal β = 0.096; 95% CI: −12.7% to +38.7%). By contrast, log prosecutorial sentence demand emerged as the dominant predictor of judicial sentence length: the adjusted log–log slope was β = 0.832 (0.8322 exact; 95% CI: 0.702 to 0.963; p < 0.001), indicating that a +10% increase in prosecutorial sentence demand was associated with a +8.25% (~8.2%) increase in judicial sentence length via exact power transformation ($1.10^{0.832156} - 1$; 95% CI: +6.92% to +9.61%, reported as +6.9% to +9.6%) and +8.3% via first-order linear elasticity approximation (0.10 × 83.2%). Among the remaining structural covariates in Table 5, aggravation or violence indicators (n = 109 present vs. 11 absent) were associated with a +22.3% difference (β = 0.201 [0.2015], 95% CI: −0.088 to 0.491, Δ% 95% CI: −8.4% to +63.4%, p = 0.173), multiple

co-defendants (n = 17 vs. 103) with a +11.2% difference (β = 0.106 [0.1061], 95% CI: −0.079 to 0.291, Δ% 95% CI: −7.6% to +33.8%, p = 0.260), and centered calendar year with a +2.6% change per year (β = 0.026 [0.0256], 95% CI: −0.034 to 0.085, Δ% 95% CI: −3.4% to +8.9%, p = 0.401). Across provincial fixed effects (relative to West Java, n = 30), Central Java (n = 41; β = −0.205 [−0.2053], Δ% = −18.6% exact [−18.5% from 3-dec β], p = 0.138), East Java (n = 31; β = −0.289 [−0.2886], Δ% = −25.1%, 95% CI: −42.2% to −2.8%, p = 0.030), DI Yogyakarta (n = 15; β = −0.114 [−0.1139], Δ% = −10.8%, p = 0.454), and DKI Jakarta (n = 3; β = 1.102 [1.1024], Δ% = +201.1% exact [+201.0% from 3-dec β], 95% CI: −43.7% to +1,511.9%, p = 0.198) showed wide intervals in small cells. In the secondary multivariable logistic regression for sentences > 540 days (n = 120; McFadden pseudo-R^2 = 0.453; Table 5), adjusted odds ratios were null for both assisted counsel (aOR = 0.85 [0.847 exact], 95% CI: 0.18–3.89, p = 0.831) and unrecorded counsel (aOR = 0.84 [0.835 exact; 0.83 two-decimal floor], 95% CI: 0.26–2.63, p = 0.758), whereas log prosecutorial demand strongly predicted above-median sentences (aOR = 34.29, 95% CI: 8.69–135.31, p < 0.001; centered year aOR = 0.92, 95% CI: 0.65–1.30, p = 0.640).

Table 5. Complete Multivariable Ordinary Least Squares (OLS) Model on Log Sentence Length with HC3 Robust Standard Errors and Complete Secondary Logistic Regression for Sentences Above the Cohort Median (n = 120 Complete Cases; OLS R^2 = 0.739; Logistic McFadden Pseudo-R^2 = 0.453).

Model & Predictor Variable (Complete-Case Subgroup n)	**Reference Category / Scale & Centering**	**Log-Scale Coefficient β (Exact / 3-Dec)**	**95% HC3 Confidence Interval for β**	**Relative Sentence Change (%) or aOR [Exact / 3-Dec β]**	**95% CI for % Change / aOR & Two-Sided P-value**
Primary OLS Model: Intercept (n = 120)	Reference covariates at centered year 2016 (year_centered = 0)	0.841 (0.8412)	−0.010 to 1.693	—	p = 0.053
Counsel Status: Assisted / Didampingi (n = 16)	vs. Unassisted / Tidak didampingi [Ref, n = 57]	−0.093 (−0.0935)	−0.350 to 0.163	−8.9% (−8.92% exact; −8.9% from 3-dec β)	−29.5% to +17.8% | p = 0.477
Counsel Status: Unrecorded / Tidak tercatat (n = 47)	vs. Unassisted / Tidak didampingi [Ref, n = 57]	0.057 (0.0569)	−0.119 to 0.233	+5.9% (+5.85% exact; +5.9% from 3-dec β)	−11.2% to +26.2% | p = 0.527
Occupation: Other or unrecorded (n = 54)	vs. Laborers and agriculture [Ref, n = 30]	0.032 (0.0317)	−0.154 to 0.218	+3.2% (+3.22% exact; +3.3% from 3-dec β)	−14.3% to +24.3% | p = 0.738
Occupation: Business and private sector (n = 36)	vs. Laborers and agriculture [Ref, n = 30]	0.096 (0.0957)	−0.135 to 0.327	+10.0% (+10.05% exact; +10.1% from 3-dec β)	−12.7% to +38.7% | p = 0.417
Log Prosecutorial Sentence Demand (n = 120)	Continuous ln(prosecution_demand_days); effect per +10% demand	0.832 (0.8322)	0.702 to 0.963	+8.3% linear (+8.25% [~8.2%] exact 1.10^β − 1)	+6.9% to +9.6% (+6.92%–+9.61%) | p < 0.001*
Aggravation or Violence Indicator (n = 109 present)	Art. 363/365 or violence phrase present vs. absent [Ref, n = 11]	0.201 (0.2015)	−0.088 to 0.491	+22.3% (+22.32% exact; +22.3% from 3-dec β)	−8.4% to +63.4% | p = 0.173

Multiple Defendants / Terdakwa jamak (n = 17 present)	Co-defendants present vs. single defendant [Ref, n = 103]	0.106 (0.1061)	−0.079 to 0.291	+11.2% (+11.19% exact; +11.2% from 3-dec β)	−7.6% to +33.8% | p = 0.260
Centered Calendar Year (year_centered = year − 2016.0)	Linear secular trend per +1 calendar year (2005–2019)	0.026 (0.0256)	−0.034 to 0.085	+2.6% (+2.59% exact; +2.6% from 3-dec β)	−3.4% to +8.9% | p = 0.401
Province: Central Java / Jawa Tengah (n = 41)	vs. West Java / Jawa Barat [Ref, n = 30]	−0.205 (−0.2053)	−0.476 to 0.066	−18.6% exact (−18.5% from 3-dec β = −0.205)	−37.9% to +6.8% | p = 0.138
Province: East Java / Jawa Timur (n = 31)	vs. West Java / Jawa Barat [Ref, n = 30]	−0.289 (−0.2886)	−0.549 to −0.028	−25.1% (−25.07% exact; −25.1% from 3-dec β)	−42.2% to −2.8% | p = 0.030*
Province: DKI Jakarta (sparse cell, n = 3)	vs. West Java / Jawa Barat [Ref, n = 30]	1.102 (1.1024)	−0.575 to 2.780	+201.1% exact (+201.0% from 3-dec β = 1.102)	−43.7% to +1,511.9% | p = 0.198
Province: DI Yogyakarta (n = 15)	vs. West Java / Jawa Barat [Ref, n = 30]	−0.114 (−0.1139)	−0.413 to 0.185	−10.8% (−10.76% exact; −10.8% from 3-dec β)	−33.8% to +20.3% | p = 0.454
Secondary Logistic Model (> 540 Days): Assisted Counsel (n = 16)	vs. Unassisted [Ref, n = 57] for sentence > 540 days	−0.166 (log-odds)	−1.715 to 1.358	aOR = 0.85 (0.847 exact)	95% HC3 CI: 0.18–3.89 (0.184–3.889) | p = 0.831
Secondary Logistic Model (> 540 Days): Unrecorded Counsel (n = 47)	vs. Unassisted [Ref, n = 57] for sentence > 540 days	−0.180 (log-odds)	−1.347 to 0.967	aOR = 0.84 (0.835 exact; 0.83 two-dec floor)	95% HC3 CI: 0.26–2.63 (0.260–2.630) | p = 0.758
Secondary Logistic Model (> 540 Days): Log Prosecutorial Demand	Continuous ln(prosecution_demand_days) per +1 log-unit	3.535 (log-odds)	2.162 to 4.908	aOR = 34.29 (34.292 exact)	95% HC3 CI: 8.69–135.31 | p < 0.001*
Secondary Logistic Model (> 540 Days): Centered Calendar Year	Linear trend per +1 year (year_centered = year − 2016.0)	−0.080 (log-odds)	−0.425 to 0.261	aOR = 0.92 (0.923 exact)	95% HC3 CI: 0.65–1.30 (0.654–1.298) | p = 0.640
Secondary Logistic Model (> 540 Days): Intercept	Baseline log-odds at reference counsel & year_centered = 0	−22.264 (log-odds)	−31.118 to −13.410	OR = 2.14×10^{-10}	95% HC3 CI: 3.06×10^{-14} to 1.50×10^{-6} | p < 0.001*

**Statistically significant at α = 0.05. Primary multivariable model estimated via ordinary least squares (OLS) on natural log-transformed sentence days among n = 120 complete-case decisions with positive sentences and available prosecutorial demands (R^2 = 0.739; 73.9% variance explained), using MacKinnon–White HC3 heteroskedasticity-consistent standard errors. Calendar year is centered at the convicted-cohort median*

(year_centered = year − 2016.0). Both exact floating-point exponentiated percentage changes [(exp(β_exact) − 1) × 100%] and 3-decimal rounded transformations [(exp(β_3dec) − 1) × 100%] are reported for complete numerical transparency. For log prosecutorial demand (slope β = 0.832156), a +10% demand increase corresponds to +8.25% (~8.2%) via exact power transformation ($1.10^{0.832156} − 1$) and +8.3% via first-order linear elasticity scaling (0.10 × 83.2%). The secondary binary logistic model for sentences strictly exceeding the cohort median (> 540 days; n = 120; McFadden pseudo-R^2 = 0.453; HC3 robust SEs) uses a parsimonious specification adjusting for counsel status, log prosecutorial demand, and centered year to avoid quasi-complete separation in sparse provincial/occupational cells. Abbreviations: aOR, adjusted odds ratio; CI, confidence interval; HC3, heteroskedasticity-consistent estimator type 3; OLS, ordinary least squares.

The multivariable forest plot of adjusted percentage sentence differences, the adjusted demand–sentence association curve, and the secondary logistic regression odds ratios for sentences exceeding the 540-day cohort median are illustrated in Figure 4 (Panel a and Panel b). As depicted in Figure 4a, the 95% HC3 confidence intervals for assisted counsel (−29.5% to +17.8%), unrecorded counsel (−11.2% to +26.2%), and occupational categories all straddle the zero-difference reference line, whereas the prosecutorial demand association (+8.3% linear / +8.25% exact per +10% demand rise) is tightly bounded. In Figure 4b, the fitted power-law curve is calibrated at the empirical complete-case medians (prosecutorial demand 720 days → judicial sentence 540 days; median sentence-to-demand ratio = 0.866 [86.6%]), while the lower forest sub-panel visually confirms the wide, null adjusted odds ratio intervals for both assisted (aOR = 0.85) and unrecorded (aOR = 0.84) representation.

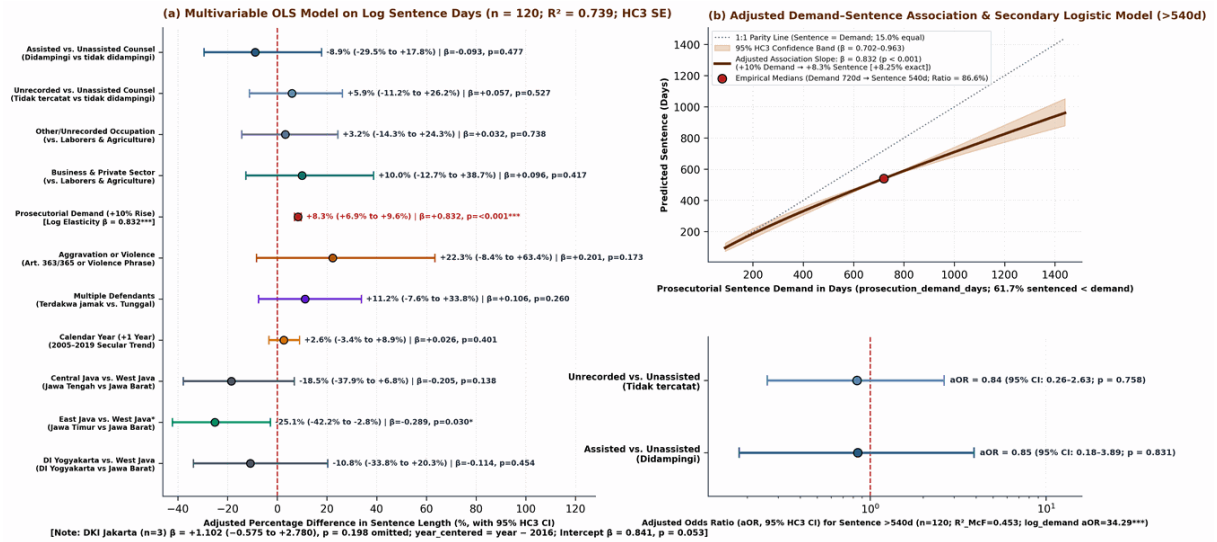

Figure 4. Multivariable ordinary least squares (OLS) regression forest plot with MacKinnon–White HC3 heteroskedasticity-consistent 95% confidence intervals showing adjusted percentage differences in sentence length across predictors in the complete-case cohort (n = 120; R^2 = 0.739; year_centered = year − 2016.0) (Panel a), and adjusted log–log association curve between prosecutorial demand and custodial sentence calibrated at the complete-case empirical medians (demand 720 days → sentence 540 days; median ratio = 86.6%; slope β = 0.832, 95% CI: 0.702–0.963, p < 0.001) paired with the secondary multivariable logistic regression forest plot for sentences exceeding the cohort median (> 540 days; McFadden pseudo-R^2 = 0.453) (Panel b).

Juridical Bifurcation, Sensitivity Cohort Robustness, and Corpus-Wide Explicit Legal Audit

Synthesizing the statutory charge breakdown with doctrinal qualification rules explains why nominally low-value theft in general district court archives bifurcates from the juridical population of statutory minor theft (pencurian ringan), as illustrated in Figure 5 (Panel a and Panel b). As mapped in the juridical bifurcation tree in Figure 5a, 0 of 148 primary cohort decisions cited Article 364 KUHP, whereas 133 of 148 (89.9%) involved statutory aggravation (Article 363 KUHP [53.4%]: nighttime residential entry, breaking/climbing, or concerted action by two or more co-defendants) or violence (Article 365 KUHP [44.6%]), disqualifying them from summary minor-theft adjudication under Article 364 KUHP and Perma Number 2 of 2012 despite their explicit monetary loss not exceeding IDR 2,500,000. Simultaneously, the structural directed acyclic graph (DAG) in Figure 5b illustrates how case severity and unmeasured case complexity act as shared upstream confounders

influencing both legal counsel recording and prosecutorial sentence demand (U → M, Y), while post-indictment prosecutorial demand acts as an intermediate node on the pathway between early defense representation and judicial sentence length (A → M → Y).

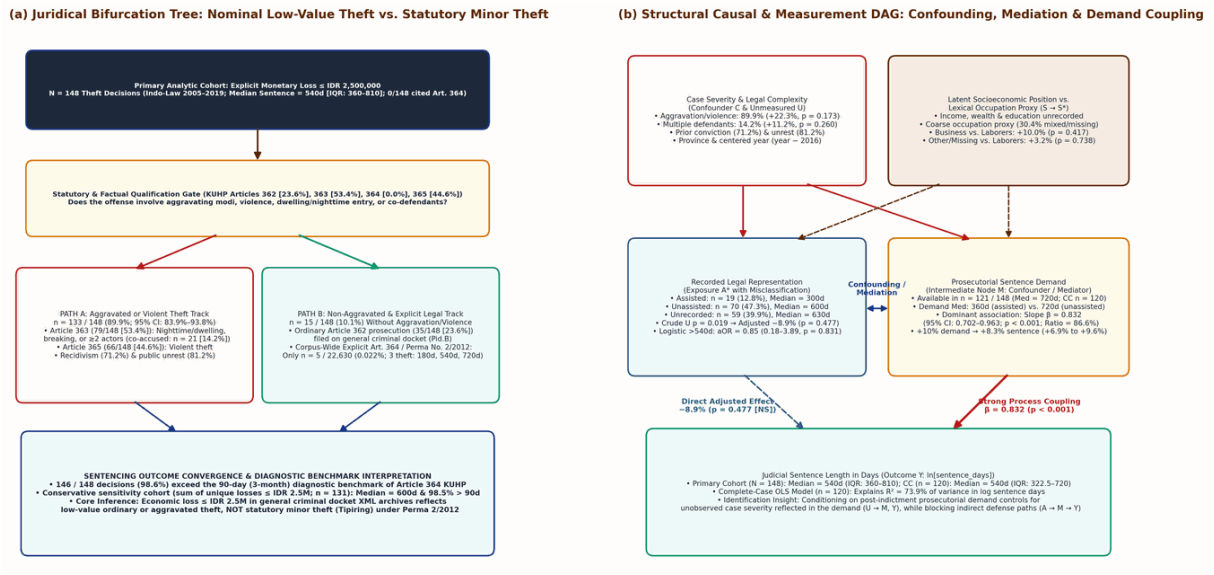

Figure 5. Juridical bifurcation tree illustrating why low-value theft decisions (explicit loss ≤ IDR 2,500,000; N = 148) predominantly follow the ordinary/aggravated prosecution track (53.4% citing Article 363, 44.6% citing Article 365, 23.6% citing Article 362, 0.0% citing Article 364; 89.9% composite aggravation/violence; 98.6% sentenced to > 90 days) rather than the statutory Article 364 / Perma No. 2/2012 summary track (Panel a), and structural causal and measurement directed acyclic graph (DAG) depicting confounding, post-indictment prosecutorial demand coupling (β = 0.832; median demand 360d assisted vs. 720d unassisted), mediation (A → M → Y), and nondifferential misclassification across the courtroom workgroup (Panel b).

Comparing the Primary Analytic Cohort (maximum contextual loss ≤ IDR 2,500,000; N = 148) against the Conservative Sensitivity Cohort (cumulative sum of all unique detected monetary values ≤ IDR 2,500,000; n = 131) confirmed high stability across operational loss definitions, as detailed in Table 6 (Panel A): the median sentence was 540 days (IQR: 360–810) in the primary cohort versus 600 days (IQR: 360–810) in the conservative cohort, and the proportion receiving custodial sentences > 90 days was 98.6% (146/148) versus 98.5% (129/131). Furthermore, executing a targeted lexical search across all N = 22,630 XML documents in the Indo-Law repository for explicit references to Article 364 KUHP (Pasal 364), minor theft (pencurian ringan), or Supreme Court Regulation Number 2 of 2012 (Perma 2/2012) identified only n = 5 relevant documents across the entire 15-year corpus (0.022% of 22,630), of which only 3 carried the source metadata label pencurian, as detailed in Table 6 (Panel B). Specifically, the five corpus-wide explicit legal hits comprised: (1) decision hash 1af98e6b… (2018, East Java; source label tidak-diketahui [unknown]; relevance basis: 'Perma 2/2012, tipiring'; verdict lepas [dismissed/exempt from punishment]; sentence = 0 days); (2) decision hash 312cd2f4… (2012, DI Yogyakarta; source label pencurian; relevance basis: 'Pasal 364'; verdict pidana [convicted]; sentence = 540 days); (3) decision hash 5fd19e7a… (2012, East Java; source label kejahatan-terhadap-keamanan-negara [crimes against state security]; relevance basis: 'Pasal 364'; verdict bebas [acquitted]; sentence = 0 days); (4) decision hash 66924472… (2017, East Java; source label pencurian; relevance basis: 'Pasal 364'; verdict pidana; sentence = 720 days); and (5) decision hash 9aa0e551… (2019, East Java; source label pencurian; relevance basis: 'Pasal 364'; verdict pidana; sentence = 180 days), as detailed in Table 6.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue X October 2026

Table 6. Sensitivity Analysis of Cohort Loss Definitions (N = 148 vs. n = 131) and Corpus-Wide Audit of Decisions Explicitly Referencing Article 364 KUHP or Perma No. 2 of 2012 (n = 5 Across N = 22,630 Decisions).

Panel / Cohort Definition or Decision Hash ID	Operational Criteria or Decision Year & Province	Sample Size or Source Corpus Label	Median Sentence or Verdict Disposition (Amar)	Sentence > 90 Days or Exact Custodial Duration	Exact Relevance Basis & Juridical Interpretation
Panel A: Primary Low-Value Cohort Definition	Maximum detected contextual monetary loss ≤ IDR 2,500,000, zero material OCR conflicts, no >threshold phrases	N = 148 (100.0% of primary)	Median: 540 days (IQR: 360–810)	146 / 148 (98.6% > 90 days)	Baseline operational cohort capturing single-figure and multi-figure low-value theft narratives (0/148 explicitly cited Art. 364)
Panel A: Conservative Sensitivity Cohort Definition	Cumulative sum of all unique detected monetary values across document ≤ IDR 2,500,000, zero material OCR conflicts	n = 131 (88.5% of primary)	Median: 600 days (IQR: 360–810)	129 / 131 (98.5% > 90 days)	Confirms that excluding 17 multi-figure decisions leaves median sentence (600 vs. 540 days) and >90-day rate (98.5% vs. 98.6%) virtually unchanged
Panel B: Explicit Legal Hit #1 — Hash 1af98e6b…	Decision Year: 2018 | Province: East Java (Jawa Timur)	Source label: unknown (tidak-diketahui)	Dismissed / exempt from punishment (lepas)	0 days (≤ 90 days)	Relevance basis: 'Perma 2/2012, tipiring' — Explicitly cites Perma No. 2/2012 and minor-offense procedure but lacks theft sub-classification label in Indo-Law metadata
Panel B: Explicit Legal Hit #2 — Hash 312cd2f4…	Decision Year: 2012 | Province: DI Yogyakarta	Source label: theft (pencurian)	Convicted / imprisonment (pidana)	540 days (> 90 days)	Relevance basis: 'Pasal 364' — Labeled theft and references Article 364 KUHP in text, yet sentenced to 540 days under ordinary/aggravated charges
Panel B: Explicit Legal Hit #3 — Hash 5fd19e7a…	Decision Year: 2012 | Province: East Java (Jawa Timur)	Source label: crimes against state security (kejahatan-terhadap-keamanan-negara)	Acquitted (bebas)	0 days (≤ 90 days)	Relevance basis: 'Pasal 364' — Demonstrates metadata mislabeling in source repository alongside non-custodial acquittal disposition
Panel B: Explicit Legal Hit #4	Decision Year: 2017 | Province: East Java (Jawa Timur)	Source label: theft (pencurian)	Convicted / imprisonment (pidana)	720 days (> 90 days)	Relevance basis: 'Pasal 364' — Illustrates rejection of defense Article 364 minor-theft arguments when

— Hash 66924472…					aggravating elements (Articles 363/365) are proven at trial
Panel B: Explicit Legal Hit #5 — Hash 9aa0e551…	Decision Year: 2019 | Province: East Java (Jawa Timur)	Source label: theft (pencurian)	Convicted / imprisonment (pidana)	180 days (> 90 days)	Relevance basis: 'Pasal 364' — Only 5 documents in all 22,630 Indo-Law XML files (0.022%; 3 labeled theft) explicitly mention Article 364, minor theft, or Perma 2/2012

Abbreviations: IDR, Indonesian Rupiah; IQR, interquartile range; KUHP, Indonesian Criminal Code; OCR, optical character recognition; Perma, Supreme Court Regulation.

The visual comparison of sensitivity cohort robustness, the case-level sentence durations and relevance bases of the five corpus-wide explicit Article 364 / Perma 2/2012 documents, and the five-stage empirical legal research roadmap for post-2026 Criminal Code reform are illustrated in Figure 6 (Panel a and Panel b). As shown in Figure 6a, both the primary cohort (N = 148) and the conservative sensitivity cohort (n = 131) exhibit virtually identical proportions of sentences exceeding 90 days (98.6% vs. 98.5%), while the three convicted theft cases among the five explicit legal hits received 180, 540, and 720 days and the two non-theft-labeled hits resulted in non-custodial dispositions (0 days). To address these structural archive boundaries as Indonesia transitions to the National Criminal Code (Law Number 1 of 2023, effective 2 January 2026), Figure 6b outlines a five-phase methodological roadmap integrating dual-annotator Pid.C/Pid.B sampling, stage-specific legal aid instrumentation, multidimensional socioeconomic linkage, quasi-experimental identification, and longitudinal post-2026 statutory transition auditing.

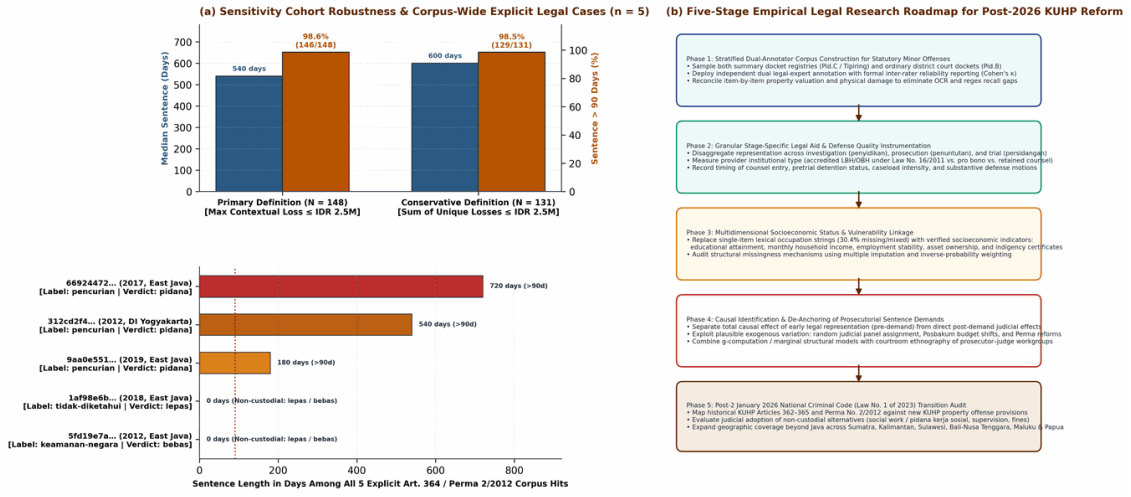

Figure 6. Sensitivity comparison between the primary operational cohort definition (maximum loss ≤ IDR 2,500,000; N = 148; median 540 days; 98.6% > 90 days) and the conservative cumulative-loss definition (sum of unique losses ≤ IDR 2,500,000; n = 131; median 600 days; 98.5% > 90 days) alongside the sentence profile and relevance basis of all 5 corpus-wide explicit Article 364 / Perma No. 2/2012 decisions (Panel a), and five-stage empirical legal research and policy roadmap for the post-2 January 2026 Indonesian Criminal Code (Law No. 1 of 2023) transition (Panel b).

DISCUSSION

The empirical findings of this observational content analysis of 148 nominally low-value theft decisions (explicit monetary loss ≤ IDR 2,500,000) extracted from the 22,630-document Indo-Law repository yield three core methodological and socio-legal insights, as synthesized in Table 7.[1–10,26] First, while unadjusted bivariate comparison revealed a two-fold sentencing gradient in which defendants with recorded legal representation

received a median custodial sentence of 300 days (IQR: 165–585) compared with 600 days (IQR: 375–787) among explicitly unrepresented defendants (Mann–Whitney U test p = 0.019), this crude association did not persist after multivariable adjustment in the complete-case cohort (n = 120; R^2 = 0.739), attenuating to a non-significant difference of −8.9% (95% HC3 CI: −29.5% to +17.8%, p = 0.477; secondary logistic aOR for > 540 days = 0.85, 95% CI: 0.18–3.89, p = 0.831).[6,7,27,28] Second, lexical defendant occupation categories—used as a coarse proxy for socioeconomic position—exhibited higher crude medians among manual laborers/agricultural workers (630 days; n = 39 [26.4%]) and unemployed defendants (630 days; n = 13 [8.8%]) than among business/private-sector workers (540 days; n = 44 [29.7%]), yet showed no independent multivariable association (+10.0%, p = 0.417 for business/private; +3.2%, p = 0.738 for other/unrecorded occupation).[8,29] Third, log prosecutorial sentence demand (median 720 days in complete cases) exhibited a strong adjusted association with judicial sentence length (median 540 days; slope β = 0.832, 95% CI: 0.702–0.963, p < 0.001; +8.3% linear / +8.25% exact sentence change per +10% demand difference), while 89.9% (133/148) of low-value theft decisions carried Article 363/365 aggravation or violence indicators (0/148 cited Article 364), resulting in 98.6% (146/148) exceeding the 90-day Article 364 diagnostic benchmark.[4,6,9–15]

Table 7. Comparative Empirical Synthesis and Five-Pillar Methodological Roadmap for Sentencing Disparity and Legal Aid Research Under the 2026 Indonesian Criminal Code (Law No. 1 of 2023).

Empirical / Policy Domain	**Key Finding in Indo-Law Cohort (2005–2019; N = 148)**	**Global & Indonesian Comparative Benchmark[1–45]**	**Identified Methodological or Institutional Limitation**	**Actionable Research & Judicial Policy Recommendation (Post-2026 KUHP Era)**
1. Nominal Low-Value vs. Juridical Minor Theft Bifurcation	89.9% (133/148) carry Article 363 (53.4%), Article 365 (44.6%), or violence indicators (0/148 cited Art. 364); 98.6% (146/148) exceed the 90-day diagnostic benchmark (median 540 days); only 5 corpus-wide decisions explicitly cite Art. 364/Perma 2/2012	Aligns with Indonesian socio-legal studies showing summary Tipiring dockets (Pid.C) are archived separately from ordinary Pid.B dockets and precluded by night/dwelling or concerted modi[1,4,11–15]	General criminal XML repositories under-index summary registries and cannot estimate national Perma No. 2/2012 compliance	Construct stratified multi-docket corpora sampling both Pid.C summary registers and Pid.B ordinary dockets with dual legal-expert annotation (Cohen's κ)[1,4,14]
2. Legal Representation Recording & Confounding Attenuation	Assisted cases (19/148; 12.8%) show shorter crude sentences than unassisted cases (70/148; 47.3%): median 300 vs. 600 days (p = 0.019), but adjusted OLS difference is −8.9% (p = 0.477) and logistic aOR = 0.85 (p = 0.831)	Converges with empirical defense-counsel literature demonstrating that binary representation indicators are confounded by charge severity and attenuated by unmeasured defense quality and caseloads[7,9,36]	39.9% (59/148) have unrecorded counsel status (nondifferential misclassification bias toward null); timing, provider type (LBH vs. retained), and quality are unmeasured	Mandate structured XML metadata fields for counsel stage (investigation, prosecution, trial) and provider accreditation under Law No. 16 of 2011^8,36
3. Prosecutorial Demand	Prosecutorial demand (median 720d in n=120; 360d assisted vs. 720d	Corroborates international courtroom	Observational log–log slope cannot isolate	Implement two-stage causal mediation / g-computation models

Alignment & Process Coupling	unassisted) is the dominant predictor of sentence length (median 540d; median ratio 86.6%): slope $\beta = 0.832$ (95% CI: 0.702–0.963, $p < 0.001$; +8.3% [+8.25% exact] sentence per +10% demand), driving $R^2 = 0.739$	workgroup scholarship showing judicial sentences closely track prosecutorial recommendations while typically applying a modest downward discount (61.7% sentenced below demand)[6,9,10,25,29]	causal cognitive anchoring from shared unmeasured case severity (U → M, Y) or post-treatment mediation of early defense advocacy (A → M → Y)	separating pre-demand legal aid effects from post-demand judicial sentencing adjustments[6,7,26]
4. Socioeconomic Proxy Granularity & Missingness	Lexical occupation groups show 630-day medians for laborers/agriculture (26.4%) and unemployed (8.8%) vs. 540 days for business/private (29.7%), but adjusted differences are null (+10.0%, $p = 0.417$; +3.2%, $p = 0.738$)	Echoes criminological focal-concerns studies warning that single-item occupational strings suffer high missingness (30.4% mixed/unrecorded; 12.8% missing <identitas>) and fail to capture income, assets, or pretrial detention[8,29,33]	Education, monthly income, household wealth, and indigency certificates (SKTM) are not systematically standardized in decision headers	Link court decision registries with standardized socioeconomic vulnerability indices and audit missing-data mechanisms via multiple imputation[7,8,26]
5. Transition to the 2023 National Criminal Code (Effective 2 Jan 2026)	Historical cohort (2005–2019) exhibits near-universal custodial sentencing (99.3% >0 days; 98.6% >90 days) despite high remorse (95.0%) and low monetary loss (≤ IDR 2.5M)	Law No. 1 of 2023 (effective 2 January 2026) and prosecutorial restorative justice regulations introduce social work punishment (pidana kerja sosial), supervision, and statutory sentencing guidelines[5,10,16–18,37–45]	Pre-2020 decisions from five Java provinces (with sparse cells such as DKI Jakarta n = 3) cannot be extrapolated to post-2026 national judicial practice across all 38 Indonesian provinces	Execute prospective longitudinal evaluations of post-2026 decisions mapping old KUHP Articles 362–365 to new KUHP provisions and non-custodial alternatives[5,10,17,38]

Abbreviations: aOR, adjusted odds ratio; CI, confidence interval; IDR, Indonesian Rupiah; KUHP, Indonesian Criminal Code; LBH, Legal Aid Institute (Lembaga Bantuan Hukum); OLS, ordinary least squares; Perma, Supreme Court Regulation; SKTM, Certificate of Financial Indigency (Surat Keterangan Tidak Mampu); XML, Extensible Markup Language.

Interpreting the attenuation of the legal representation gradient from a crude 50.0% median difference (300 vs. 600 days, $p = 0.019$) to an adjusted −8.9% difference ($p = 0.477$) requires distinguishing pre-trial confounding, post-indictment mediation, and administrative measurement error (Figure 5b).[6,7,9,26,28] In observational judicial archives, legal representation is non-randomly assigned, and our complete-case stratification revealed that assisted defendants entered sentencing with a median prosecutorial demand of 360 days (IQR: 232.5–360; n = 16) compared with 720 days (IQR: 360–720; n = 57) among unassisted defendants.[7,29,36] Crucially, this

attenuation does not prove that legal aid has zero causal impact on sentencing outcomes.[7,36] First, the categorical variable counsel_status captures only whether the written judgment explicitly records counsel; it cannot disaggregate the timing of representation (investigation [penyidikan], prosecution [penuntutan], or trial [persidangan]), institutional provider type (accredited legal aid organizations [Lembaga Bantuan Hukum / OBH] under Law Number 16 of 2011 vs. court-appointed pro bono counsel vs. privately retained advocates), or defense quality.[7,36] Second, 39.9% of the primary cohort (59/148 decisions; median sentence 630 days) had unrecorded counsel status (tidak tercatat), and nondifferential misclassification between represented and unrepresented cases biases adjusted estimates toward the null.[26] Third, because prosecutors formulate their sentencing demands near the end of trial after defense cross-examination and mitigation advocacy, prosecutorial demand lies on the causal pathway between early defense representation and the final judicial verdict (A → M → Y in Figure 5b): conditioning on prosecutorial demand controls for baseline offense severity reflected in the demand, but blocks any indirect protective effect of counsel mediated through lower prosecutorial demands.[6,7,9]

Likewise, the strong adjusted association between log prosecutorial demand and log judicial sentence length ($\beta = 0.832$, 95% CI: 0.702–0.963, $p < 0.001$; logistic aOR = 34.29 for > 540 days) must be interpreted strictly as an observational process association rather than a proven causal cognitive anchoring effect.[6,9,10,25] Across the complete-case cohort (n = 120), prosecutorial demands had a median of 720 days (IQR: 360–810) whereas imposed custodial sentences had a median of 540 days (IQR: 322.5–720), yielding a median sentence-to-demand ratio of 0.866 (86.6%, IQR: 69.2%–100.0%), with 61.7% (74/120) of defendants sentenced below the prosecutor's demand, 15.0% (18/120) sentenced equal to the demand, and 23.3% (28/120) having an imposed sentence exceeding the single-clause extracted demand figure. Because both the prosecutor's demand and the judicial sentence respond to sharedcase features that are only partially captured by binary indicators—such as detailed evidentiary strength, criminal record nuances (71.2% recidivism in structured considerations), weapon lethality, and local court norms (U → M, Y)—an observational regression cannot separate psychological anchoring from shared institutional evaluation of offense gravity.[6,9,24,25]

The null multivariable findings across collapsed defendant occupation groups (+10.0% [+10.05% exact] for business/private sector, $p = 0.417$; +3.2% [+3.22% exact] for other/unrecorded occupation, $p = 0.738$ relative to laborers and agriculture) must similarly be qualified by severe measurement and sample-size constraints.[8,16,26,33] Single-word occupational strings in court headers conflate informal street traders with salaried private-sector employees, 30.4% of cases (45/148) had mixed or unrecorded occupations (including 12.8% [19/148] lacking a segmented <identitas> section), and minor categories (students n = 4, homemakers n = 2, other formal n = 1) were too sparse for standalone estimation.[1,8] Direct socioeconomic measures—monthly income, household wealth, education, indigency certificates (Surat Keterangan Tidak Mampu [SKTM]), ability to pay victim restitution (documented in only 8.8% [7/80] of structured considerations), and pretrial detention duration—are unrecorded in Indo-Law XML metadata.[1,8,33,34]

Doctrinally, our finding that 0 of 148 primary cohort cases cited Article 364 KUHP while 89.9% (133/148) carried Article 363 (53.4%) or Article 365 (44.6%) aggravation or violence indicators—and 98.6% (146/148) exceeded the 90-day benchmark—demonstrates that nominal loss ≤ IDR 2,500,000 in general district court XML archives represents low-value ordinary or aggravated theft rather than statutory minor theft (pencurian ringan).[1,4,11–15] Furthermore, only 5 documents across all 22,630 Indo-Law XML files (0.022%; 3 labeled pencurian) explicitly referenced Article 364 KUHP or Perma Number 2 of 2012, confirming that summary minor-offense adjudications (acara pemeriksaan cepat / Pid.C) are archived in separate simplified registers omitted from general Pid.B XML repositories.[1,2,4]

Five key limitations bound the generalizability of this study.[26] First, the analytic sample is modest (N = 148; n = 120 complete cases), restricted to five provinces in Java, and contains sparse cells in specific provinces (DKI Jakarta n = 3) and occupations (n ≤ 4), precluding national extrapolation across all 38 Indonesian provinces. Second, the corpus spans 2005–2019 and predates both modern restorative justice regulations (Prosecutor's Regulation No. 15 of 2020; Police Regulation No. 8 of 2021)[37–45] and the new Indonesian Criminal Code (Law No. 1 of 2023, effective 2 January 2026).[5] Third, Indo-Law is a curated research sample rather than a complete census of Supreme Court decisions.[1] Fourth, despite auditing 8,828 digit–word pairs and manually verifying 40 sampled decisions (random_state = 20261022), rule-based regex extraction remains vulnerable to residual OCR noise and uncaptured values in unstructured narrative sections. Fifth, substantial missingness in recorded counsel

status (39.9% unrecorded), prosecutorial demands (18.2% missing), and structured judicial considerations (45.9% unsegmented)—even though sentence lengths did not differ significantly between complete and incomplete subsets (Mann–Whitney p = 0.249 and p = 0.890)—could introduce selection or measurement bias if missingness correlates with unmeasured confounders.

CONCLUSIONS

In this observational quantitative content analysis of 148 nominally low-value theft decisions (explicit monetary loss ≤ IDR 2,500,000) extracted from the 22,630-document Indo-Law XML corpus across five provinces in Java (2005–2019), recorded legal representation was associated with shorter custodial sentences in unadjusted comparisons (median 300 days vs. 600 days among unassisted defendants; Mann–Whitney U test p = 0.019), where assisted defendants also faced lower prosecutorial demands (median 360 vs. 720 days), but showed no independent association after multivariable adjustment for prosecutorial demand, aggravated or violent offense characteristics, defendant multiplicity, centered decision year, and province (adjusted OLS difference = −8.9%, 95% HC3 CI: −29.5% to +17.8%, p = 0.477; logistic aOR for > 540 days = 0.85, 95% CI: 0.18–3.89, p = 0.831). Collapsed defendant occupation categories similarly exhibited no independent multivariable association with sentence length, whereas log prosecutorial sentence demand (median 720 days vs. 540 days sentence median; median ratio 86.6%) was the dominant statistical predictor of judicial sentencing (slope β = 0.832, 95% CI: 0.702–0.963, p < 0.001; +8.3% linear / +8.25% exact sentence duration per +10% demand difference; model R^2 = 0.739). Because 0 of 148 primary cohort decisions cited Article 364 KUHP while 89.9% (133/148) carried Article 363/365 aggravation or violence indicators, 98.6% (146/148) exceeded the 90-day Article 364 diagnostic benchmark, and only 5 decisions across the entire 22,630-document corpus explicitly referenced Article 364 KUHP or Perma Number 2 of 2012, findings must be interpreted strictly as an observational profile of nominal low-value theft in general court dockets rather than an estimate of statutory minor theft or proof of causal anchoring or the absence of socioeconomic bias.

Declarations

Ethics Approval and Privacy Protection: This study analyzed publicly accessible, de-identified secondary court judgment XML files from the open-access Indo-Law research repository (commit 67340338adbc63021630947d905627eef0d7e43f). All analytic outputs were restricted to cryptographic case hashes, aggregate metadata, and statistical model estimates, with zero personal identification numbers (NIP), defendant names, or residential addresses retained. Funding: The authors received no external financial support for the research, authorship, or publication of this article. Competing Interests: The authors declare no competing financial or non-financial interests. Data and Code Availability: The primary Indo-Law XML corpus is publicly accessible via the IR NLP Lab CSUI repository (https://github.com/ir-nlp-csui/indo-law), and all rule-based extraction scripts, de-identified cohort tables, and statistical replication files are archived in the study repository. Author Contributions: SLO, HPP, HU, and DS conceived and designed the study; EM, CN, and AE executed the rule-based XML parsing, OCR conflict reconciliation, and econometric modeling; YMP, FD, and AH performed the doctrinal legal verification, sensitivity audits, and 300 DPI visualization design; all authors (SLO, HPP, HU, DS, EM, CN, AE, YMP, FD, and AH) drafted, critically revised, and approved the final manuscript.

REFERENCES

1. Nuranti EQ, Yulianti E, Husin HS. Predicting the Category and the Length of Punishment in Indonesian Courts Based on Previous Court Decision Documents. Computers. 2022;11(6):88. https://doi.org/10.3390/computers11060088
2. Cahyawijaya S, Lovenia H, Aji AF, Winata GI, Wilie B, Mahendra R, et al. NusaCrowd: Open Source Initiative for Indonesian NLP Resources. In: Findings of the Association for Computational Linguistics: ACL 2023. Toronto: Association for Computational Linguistics; 2023. p. 13784–13818. https://doi.org/10.18653/v1/2023.findings-acl.868
3. Nasution N, Suprapto S. Court Decision Prediction Model Using Natural Language Processing and Random Forest. IJCCS (Indones J Comput Cybern Syst). 2025;19(3):305–316. https://doi.org/10.22146/ijccs.108377

4. Faradilah F. Analisis Hukum terhadap Peraturan Mahkamah Agung No. 2 Tahun 2012 Tentang Penyesuaian Batasan Tindak Pidana Ringan dan Jumlah Denda dalam Kitab Undang-Undang Hukum Pidana: Studi Kasus Putusan Nomor 1/Pid.C/2021/PN.Mjn. Eksekusi J Ilmu Huk Adm Negara. 2026;4(3):511–531. https://doi.org/10.55606/eksekusi.v4i3.2983
5. Butt S. Indonesia's new Criminal Code: indigenising and democratising Indonesian criminal law? Griffith Law Rev. 2023;32(2):190–214. https://doi.org/10.1080/10383441.2023.2243772
6. Plesničar MM. The challenges of being imperfect: how do judges and prosecutors deal with sentencing disparity. Front Sociol. 2024;9:1488786. https://doi.org/10.3389/fsoc.2024.1488786
7. Gottlieb A, Arnold K. The Effect of Public Defender and Support Staff Caseloads on Incarceration Outcomes for Felony Defendants. J Soc Soc Work Res. 2021;12(3):569–592. https://doi.org/10.1086/712924
8. D'Amato C, Holmes B, Feldmeyer B. Economic Competition and Racial/Ethnic Disparities in Sentencing: A Test of Economic Threat Perspective. Soc Sci. 2021;10(6):206. https://doi.org/10.3390/socsci10060206
9. Gazal-Ayal O, Elyoseph Z, Solomon A. Evaluating Large Language Models as Judicial Decision-Makers. Justice Q. 2026;43(5):1184–1219. https://doi.org/10.1080/07418825.2026.2618254
10. Santosa DGG, Dewi E, Fardiansyah AI, Putri DK. Assessing the Implementation and Impact of Corruption Sentencing Guidelines in Indonesia: Comparative Perspectives from the United States, England, and Wales. UUM J Leg Stud. 2026;17(2):48–67. https://doi.org/10.32890/uumjls2026.17.2.4
11. Kamilatun, Fadhilah N. Implementasi PERMA Nomor 2 Tahun 2012 tentang penyesuaian batasan tindak pidana ringan dan jumlah denda dalam KUHP terhadap pelaku pencurian ringan. J Huk Legalita. 2022;4(1):104–114. https://doi.org/10.47637/legalita.v4i1.553
12. Karsono B. Kedudukan Hukum Perma Nomor 2 Tahun 2012 dalam Mengefektivkan Pidana Denda pada Perkara Pencurian Ringan. KRTHA Bhayangkara. 2023;17(3):555–566. https://doi.org/10.31599/krtha.v17i3.788
13. Utami FR, Hamsir. Penyelesaian tindak pidana pencurian ringan berdasarkan penyesuaian batasan tindak pidana ringan dan jumlah denda. Alauddin Law Dev J. 2021;3(1):171–177. https://doi.org/10.24252/aldev.v3i1.15093
14. Berutu SP, Harefa RN, Manalu EP, Sebayang AGR, Nainggolan SDP. Analisis yuridis Putusan Nomor 152/Pid.C/2014/PN.Rap dikaitkan dengan Perma tentang tindak pidana ringan Nomor 2 Tahun 2012. J Rev Pendidik Pengajaran. 2024;7(2):4691–4702. https://doi.org/10.31004/jrpp.v7i2.27275
15. Hamdiyah. Analisis unsur-unsur tindak pidana pencurian: Tinjauan hukum. J Tahqiqa. 2024;18(1):98–108. https://doi.org/10.61393/tahqiqa.v18i1.216
16. Zulyadi R, Hossain MB. Alternative Criminal Punishments for the Settlement of Misdemeanor in a Social Justice Perspective. Law Reform. 2022;18(1):43–57. https://doi.org/10.14710/lr.v18i1.44712
17. Hamja. Pembaruan kebijakan pemidanaan dalam rangka mengatasi faktor penyebab overcrowding lembaga pemasyarakatan di Indonesia. Masalah-Masalah Hukum. 2024;53(3):261–270. https://doi.org/10.14710/mmh.53.3.2024.261-270
18. Irmawanti ND, Arief BN. Urgensi tujuan dan pedoman pemidanaan dalam rangka pembaharuan sistem pemidanaan hukum pidana. J Pembangunan Huk Indones. 2021;3(2):217–227. https://doi.org/10.14710/jphi.v3i2.217-227
19. Shen Y, Wei H, Tian X. TA-LJP: Term-Aware Legal Judgment Prediction. Information. 2026;17(1):17. https://doi.org/10.3390/info17010017
20. Chen Y, Zhu X, Zeng Z, Wang P, Zhu X. A legal judgment prediction model based on knowledge fusion and dependency masking. PLoS One. 2026;21(1):e0340717. https://doi.org/10.1371/journal.pone.0340717
21. Lyu Y, Wang Z, Ren Z, Ren P, Chen Z, Liu X, et al. Improving legal judgment prediction through reinforced criminal element extraction. Inf Process Manag. 2022;59(1):102780. https://doi.org/10.1016/j.ipm.2021.102780
22. Mumcuoğlu E, Öztürk CE, Ozaktas HM, Koç A. Natural language processing in law: Prediction of outcomes in the higher courts of Turkey. Inf Process Manag. 2021;58(5):102684. https://doi.org/10.1016/j.ipm.2021.102684

23. Dina NZ, Ravana SD, Idris N. Legal Judgment Prediction using Natural Language Processing and Machine Learning Methods: A Systematic Literature Review. SAGE Open. 2025;15(2):21582440251329663. https://doi.org/10.1177/21582440251329663
24. Fine A, Berthelot ER, Marsh S. Public Perceptions of Judges' Use of AI Tools in Courtroom Decision-Making: An Examination of Legitimacy, Fairness, Trust, and Procedural Justice. Behav Sci (Basel). 2025;15(4):476. https://doi.org/10.3390/bs15040476
25. Kiejnich-Kruk K, Twardawa M, Formanowicz P. Towards automated recommendations for drunk driving penalties in Poland—A case study analysis in selected court. Found Comput Decis Sci. 2023;48(4):425–451. https://doi.org/10.2478/fcds-2023-0019
26. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12(10):e1001885. https://doi.org/10.1371/journal.pmed.1001885
27. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat. 1947;18(1):50–60. https://doi.org/10.1214/aoms/1177730491
28. MacKinnon JG, White H. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. J Econom. 1985;29(3):305–325. https://doi.org/10.1016/0304-4076(85)90158-7
29. Cole K, Rydberg J, Cassidy M. Going off the grid: Assessing sentencing departures for individuals convicted of sexual offenses. Crime Delinq. 2021;67(11):1676–1705. https://doi.org/10.1177/0011128720977451
30. Hester R. Punishing for the past (sometimes): Judicial perspectives on criminal history enhancements. Prison J. 2021;101(4):443–465. https://doi.org/10.1177/00328855211029663
31. Putri AZ, Munandar TI, Haryadi. Faktor-faktor yang memengaruhi belum optimalnya pelaksanaan restorative justice di Polres Muaro Jambi. PAMPAS J Crim Law. 2024;5(3):297–314. https://doi.org/10.22437/pampas.v5i3.36944
32. Singh P. Determination of sentences in India: Policy and practice. Int Ann Criminol. 2023;61(3–4):314–327. https://doi.org/10.1017/cri.2023.20
33. Gunawan TJ, Sholehuddin M. Nonequivalence between imprisonment and fines in United States sentencing guidelines. J Law Sustain Dev. 2023;11(9):e512. https://doi.org/10.55908/sdgs.v11i9.512
34. Suzuki M, Jenkins T. Apology–forgiveness cycle in restorative justice, but how? Int Rev Victimol. 2023;29(2):259–276. https://doi.org/10.1177/02697580221079994
35. Pali B, Maglione G. Discursive representations of restorative justice in international policies. Eur J Criminol. 2023;20(2):507–527. https://doi.org/10.1177/14773708211013025
36. Ajah BO, Chinweze UC, Ajah IA, Onyejegbu DC, Obiwulu AC, Onwuama EM, et al. Behind Bars But Not Sentenced: The Role of Computerized Central Repository in Addressing Awaiting-Trial Problems in Ebonyi State, Nigeria. SAGE Open. 2022;12(1):21582440221079822. https://doi.org/10.1177/21582440221079822
37. Mulyono GP, Arief BN, Soponyono E. Progressionism restorative justice policies in achieving rehabilitative criminal justice. Indones J Crim Law Stud. 2024;9(1):117–138. https://doi.org/10.15294/ijcls.v9i1.36420
38. Wulandari C. Restorative justice-based criminal case resolution: A study at restorative justice houses in Yogyakarta. Indones J Crim Law Stud. 2025;10(1):239–266. https://doi.org/10.15294/ijcls.v10i1.24071
39. Wulandari C, Pujirahayu EW, Hiariej EOS, Hassan MS, Kambuno JA. Penal Mediation: Criminal Case Settlement Process based on the Local Customary Wisdom of Dayak Ngaju. Lex Scientia Law Rev. 2022;6(1):69–92. https://doi.org/10.15294/lesrev.v6i1.54896
40. Rochaeti N, Prasetyo MH, Park JH. Implementing of Restorative Justice to Build the Criminal Justice System in Indonesia: A Study of the Batak Toba Justice System. Law Reform. 2023;19(2):221–247. https://doi.org/10.14710/lr.v19i2.53184
41. Rochaeti N, Prasetyo MH, Rozah U, Park J. A Restorative Justice System in Indonesia: A Close View from the Indigenous Peoples' Practices. Sriwijaya Law Rev. 2023;7(1):87–104. https://doi.org/10.28946/slrev.Vol7.Iss1.1919.pp87-104
42. Syahrin A, Ekaputra M, Din M. Restorative justice in Indonesian criminal law: Integrating Pancasila values in police discretion practices. Int J Innov Res Sci Stud. 2025;8(2):1978–1985. https://doi.org/10.53894/ijirss.v8i2.5589

43. Harsya RMK, Triyantoro A. Analisis yuridis penerapan restorative justice dalam tindak pidana ringan di Indonesia. Sanskara Huk HAM. 2025;3(3):132–140. https://doi.org/10.58812/shh.v3i03.541
44. Sucipto H, Suaedi F, Setijaningrum E, Amiati M, Nasution RS. Transforming Public Trust in Restorative Justice: An Islamic and Social Law Perspective on the Prosecutor's Role in the Contemporary Era. MILRev Metro Islam Law Rev. 2024;3(2):364–387. https://doi.org/10.32332/milrev.v3i2.9938
45. Reza RA, Akli Z, Afrizal TY. Pelaksanaan restorative justice dalam perkara tindak pidana pencurian ringan menurut Qanun Aceh Nomor 9 Tahun 2008 tentang Pembinaan Kehidupan Adat dan Adat Istiadat (Studi penelitian Gampong Blang Talon Kecamatan Kuta Makmur Kabupaten Aceh Utara). J Ilm Mhs Fak Huk Univ Malikussaleh. 2024;6(4):1–12. https://doi.org/10.29103/jimfh.v6i4.12926