

LAPORAN AKHIR
PENERAPAN ASSOCIATION RULE MINING DAN CLUSTERING
UNTUK MENGENAL POLA KEMUNCULAN BERSAMA
PENGADUAN DAN TIPOLOGI WILAYAH PADA DATA
NYC 311 SERVICE REQUESTS

Diajukan untuk memenuhi tugas mata kuliah

Data Science



Disusun dan diajukan oleh:

IKRAR GEMPUR TIRANI

D121231015

DEPARTEMEN TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS HASANUDDIN

GOWA

2026

ABSTRAK

IKRAR GEMPUR TIRANI. *Penerapan Association Rule Mining dan Clustering untuk Mengkaji Pola Kemunculan Bersama Pengaduan dan Tipologi Wilayah pada Data NYC 311 Service Requests.*

Data pengaduan warga *NYC 311 Service Requests* berukuran sangat besar namun jarang ditambang untuk menemukan pola kondisi bangunan pada tingkat individual. Penelitian ini bertujuan menemukan dan memvalidasi pola tersembunyi pada data tersebut menggunakan teknik penambangan data deskriptif. Metode yang digunakan adalah penambangan aturan asosiasi dengan algoritma *FP-Growth* untuk menemukan kemunculan-bersama jenis pengaduan pada tiap bangunan, serta pengelompokan *K-means* untuk menemukan tipologi wilayah. Pola yang ditemukan kemudian divalidasi secara independen menggunakan catatan pelanggaran perumahan resmi melalui desain dua jendela waktu untuk memisahkan sinyal nyata dari kemungkinan sirkularitas administratif, dan dikendalikan terhadap ukuran bangunan melalui stratifikasi. Hasil menunjukkan adanya sindrom pemeliharaan interior, yaitu enam jenis pengaduan yang muncul bersama melebihi kebetulan dan berasosiasi dengan tingkat pelanggaran terverifikasi yang lebih tinggi per unit hunian. Asosiasi ini bertahan di setiap pita ukuran dan pada jendela waktu yang mendahului pengaduan, sehingga konsisten dengan interpretasi pengabaian kronis yang nyata. Beban pengabaian ini ditemukan tersebar dari sisi pemilik, namun menumpuk pada stok perumahan terjangkau. Pada tingkat wilayah, ditemukan tiga tipe karakter pengaduan yang bergradasi dan sangat lengket sepanjang waktu, serta tidak menunjukkan korelasi berarti dengan perubahan sosio-ekonomi. Penelitian menyimpulkan bahwa data 311 memuat sinyal terverifikasi yang konsisten dengan kondisi pengabaian bangunan yang bersifat kronis.

Kata Kunci: penambangan data, aturan asosiasi, pengelompokan, NYC 311, pengabaian bangunan

RINGKASAN EKSEKUTIF

Penelitian ini menjawab satu pertanyaan yang belum tergarap dalam literatur data pengaduan kota, yaitu apakah pola kemunculan bersama jenis pengaduan pada satu bangunan dapat berfungsi sebagai sidik jari pengabaian, dan apakah karakter pengaduan antarwilayah membentuk tipologi yang stabil. Penelitian terdahulu umumnya bekerja pada tingkat wilayah dan tidak merancang validasi yang mampu memisahkan sinyal nyata dari sirkularitas administratif. Celah inilah yang dijawab.

Data utama adalah *NYC 311 Service Requests* sebanyak 21,3 juta catatan, yang setelah pembersihan dan penyaringan menjadi 18,85 juta catatan dan dibentuk menjadi 764.114 keranjang bangunan. Data ini diperkaya dengan data properti *PLUTO*, divalidasi dengan catatan pelanggaran *HPD*, dan dikontekstualisasikan dengan data sensus *ACS*. Metode yang dipakai adalah penambangan aturan asosiasi *FP-Growth* dan pengelompokan *K-means*, dengan pengendalian variabel pengganggu melalui stratifikasi serta desain validasi dua jendela waktu (konkuren dan prior).

Temuan utama mencakup enam hal. Pertama, terbentuk sindrom pemeliharaan interior enam item, dengan total 95.902 kombinasi sering yang ditemukan, dan bangunan cenderung memuat keenam item secara lengkap. Kedua, kemunculan sindrom mengikuti dua pola temporal, yaitu ledakan pelaporan jangka pendek dan akumulasi multi-tahun. Ketiga, wilayah kota terbagi menjadi tiga tipologi pengaduan yang bergradasi (interior-perumahan, campuran-komersial, dan kendaraan-luar-ruang). Keempat, beban sindrom menumpuk pada stok perumahan terjangkau (HDFC) dengan selisih 16 sampai 27 poin persen di atas kepemilikan privat pada ukuran setara. Kelima, pergeseran karakter pengaduan tidak menunjukkan korelasi berarti dengan perubahan sosio-ekonomi (hasil nol). Keenam, sebuah temuan mengenai spesifisitas keparahan sengaja dibatalkan karena tidak bertahan pada jendela independen, sebagai bukti kedisiplinan analisis.

Keterbatasan terbesar adalah bahwa data 311 merekam pelaporan, bukan kondisi sebenarnya, sehingga seluruh temuan harus dibaca sebagai pernyataan tentang apa yang dilaporkan dan tervalidasi, bukan tentang seluruh kondisi fisik yang ada. Meskipun demikian, karena pola bertahan pada jendela yang mendahului pengaduan dan terverifikasi oleh inspeksi independen, implikasi praktisnya jelas, yaitu data 311 berpotensi menjadi instrumen diagnostik untuk penentuan prioritas inspeksi perumahan, dengan catatan bahwa ia mencerminkan kondisi yang telah terjadi dan bukan meramalkan masa depan.

Daftar Isi

ABSTRAK	i
RINGKASAN EKSEKUTIF	ii
GLOSARIUM	vii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Analisis Kesenjangan Literatur	2
1.3 Rumusan Masalah	2
1.4 Kriteria Keberhasilan	3
1.5 Tujuan Penelitian	3
1.6 Manfaat Penelitian	3
1.7 Ruang Lingkup	3
BAB II DATA PROVENANCE DAN AUDIT	5
2.1 Asal-Usul Data dan Rantai Pengolahan	5
2.2 Kamus Data	6
2.3 Analisis Eksploratori Data	7
2.4 Analisis Mekanisme Data Hilang	9
2.5 Pernyataan Keterwakilan Data	10
BAB III TINJAUAN PUSTAKA	11
3.1 Data 311 sebagai Penanda Kondisi Perkotaan	11
3.2 Penambahan Aturan Asosiasi	11
3.3 Pengelompokan dengan K-Means	12
3.4 Pendekatan CRISP-DM	13
3.5 Prinsip Penemuan, Bukan Konfirmasi	13
BAB IV METODOLOGI PENELITIAN	14
4.1 Jenis dan Sumber Data	14
4.2 Alur Pengolahan Data	14
4.3 Persiapan dan Pembersihan Data	15
4.4 Justifikasi Pemilihan Algoritma	16
4.5 Instrumen Sindrom Pemeliharaan Interior	16
4.6 Spesifikasi Parameter dan Analisis Sensitivitas	17
4.7 Desain Validasi Dua Jendela	18
4.8 Definisi Baseline	18
4.9 Teknik Pengendalian Ukuran dan Analisis Wilayah	19
BAB V HASIL	20
5.1 Catatan Pembacaan	20
5.2 Distribusi Sindrom (RQ1)	20
5.3 Struktur Keterkaitan Antarpengaduan	20
5.4 Hubungan Dosis-Respons dengan Pelanggaran	21
5.5 Hasil Spesifisitas Keparahan	22
5.6 Urutan dan Kecepatan Temporal	23
5.7 Konsentrasi Kepemilikan	24
5.8 Sindrom menurut Kategori Pemilik	24

5.9	Tipologi Wilayah (RQ2).....	25
5.10	Keterkaitan Lintas-Skala.....	26
5.11	Kestabilan dan Pergeseran Wilayah.....	27
5.12	Keselarasan dengan Sosio-Ekonomi.....	27
BAB VI PEMBAHASAN KRITIS.....		29
6.1	Interpretasi Temuan Utama.....	29
6.2	Uji Ketangguhan Interpretasi.....	29
6.3	Pemeriksaan Variabel Pengganggu.....	30
6.4	Ukuran Efek versus Signifikansi Statistik.....	31
6.5	Batas Generalisasi.....	31
BAB VII PENUTUP.....		32
7.1	Kesimpulan.....	32
7.2	Keterbatasan.....	33
7.3	Saran.....	33
LAMPIRAN.....		34
Lampiran A — Spesifikasi Lingkungan Komputasi.....		34
Lampiran B — Log Keputusan Preprocessing.....		34
Lampiran C — Tabel Hasil Lengkap.....		34
Lampiran D — Log Eksperimen yang Tidak Dipilih.....		35
DAFTAR PUSTAKA.....		36

Daftar Tabel

Tabel 1	Kriteria keberhasilan yang ditetapkan sebelum analisis.....	3
Tabel 2	Kamus data variabel kunci NYC 311 Service Requests.....	6
Tabel 3	Variabel kunci dan limitasi sumber data pendukung.....	7
Tabel 4	Empat sumber data dan perannya dalam penelitian.....	14
Tabel 5	Perbandingan tiga algoritma kandidat penambangan asosiasi.....	16
Tabel 6	Analisis sensitivitas parameter min_support penambangan asosiasi ..	17
Tabel 7	Ringkasan rasio pelanggaran sindrom-6 terhadap sindrom-0 per pita ukuran.....	22
Tabel 8	Pemeriksaan variabel pengganggu dan sisa kekhawatiran	30
Tabel 9	Batas generalisasi temuan	31

Daftar Gambar

Gambar 1	Distribusi dua puluh jenis pengaduan teratas berdasarkan volume .	7
Gambar 2	Distribusi temporal volume pengaduan bulanan periode 2020–2024	8
Gambar 3	Distribusi geografis densitas pengaduan antar-borough	8
Gambar 4	Distribusi jumlah jenis pengaduan per keranjang bangunan sebelum penyaringan sindrom	9
Gambar 5	Alur pengolahan data dari 21,3 juta pengaduan hingga temuan tervalidasi	15
Gambar 6	Desain dua jendela untuk memisahkan sinyal nyata dari sirkularitas administratif	18
Gambar 7	Jumlah bangunan menurut banyaknya item sindrom interior yang dilaporkan	20
Gambar 8	Struktur ko-okurensi keluhan kerusakan interior berdasarkan nilai lift	21
Gambar 9	Tingkat pelanggaran HPD per unit menurut jumlah item sindrom di setiap pita ukuran dan kedua jendela waktu	22
Gambar 10	Rasio pelanggaran per kelas keparahan pada jendela konkuren dan jendela prior	22
Gambar 11	Item yang pertama dilaporkan dinyatakan sebagai rasio terhadap ekspektasi volume	23
Gambar 12	Distribusi kecepatan akumulasi sindrom yang bersifat dua puncak	23
Gambar 13	Konsentrasi kepemilikan gedung terabaikan dibandingkan baseline kepemilikan umum	24
Gambar 14	Persen gedung ber-sindrom per kategori pemilik dengan kontrol ukuran gedung	25
Gambar 15	Tipologi wilayah berupa item paling khas per cluster dan komposisi borough	25
Gambar 16	Wujud tiga tipe wilayah dalam ruang PCA	26
Gambar 17	Konsentrasi gedung ber-sindrom RQ1 menurut tipe cluster wilayah RQ2	26
Gambar 18	Magnitudo dibandingkan konsistensi arah pergeseran komposisi keluhan antarwilayah	27
Gambar 19	Hasil uji korelasi pergeseran karakter keluhan dengan perubahan sosio-ekonomi	28
Gambar 20	Sintesis temuan mengenai apa yang dikatakan data 311 tentang pengabaian bangunan	32

GLOSARIUM

Istilah	Definisi
<i>Support</i>	Proporsi keranjang yang memuat suatu item atau kombinasi item terhadap seluruh keranjang.
<i>Confidence</i>	Probabilitas suatu item muncul dengan syarat item lain sudah ada dalam keranjang yang sama.
<i>Lift</i>	Rasio antara <i>support</i> gabungan dua item terhadap perkalian <i>support</i> masing-masing; nilai di atas satu menandakan asosiasi melebihi kebetulan.
<i>FP-Growth</i>	Algoritma penambangan pola sering yang memakai struktur pohon (<i>FP-tree</i>) tanpa membangkitkan kandidat kombinasi secara eksplisit.
<i>Frequent itemset</i>	Kombinasi item yang muncul bersama dengan frekuensi di atas ambang <i>support</i> minimum.
Sindrom pemeliharaan interior	Instrumen enam item pengaduan (Pemanas/Air Panas, Kondisi Tidak Higienis, Pipa, Cat/Plester, Kebocoran Air, Pintu/Jendela) yang muncul bersama sebagai penanda pengabaian.
<i>Silhouette score</i>	Ukuran kualitas pengelompokan bernilai -1 hingga 1 yang menilai seberapa rapat objek dengan kelompoknya dibanding kelompok tetangga.
<i>K-means</i>	Algoritma pengelompokan yang membagi objek ke dalam k kelompok dengan meminimalkan jarak ke pusat kelompok.
<i>Pooled rate</i>	Total pelanggaran dibagi total unit pada satu kelompok, sebagai ukuran intensitas yang tahan terhadap prevalensi rendah.
MCAR	<i>Missing Completely At Random</i> , data hilang sepenuhnya acak dan tidak bergantung variabel mana pun.
MAR	<i>Missing At Random</i> , data hilang bergantung pada variabel lain yang terobservasi.
MNAR	<i>Missing Not At Random</i> , data hilang bergantung pada nilai variabel itu sendiri yang tidak terobservasi.
Jendela konkuren	Periode pelanggaran HPD sejak 2020 yang sezaman dengan periode pengaduan, sehingga rentan sirkularitas.
Jendela prior	Periode pelanggaran HPD 2015–2019 yang mendahului pengaduan, sehingga bebas dari sirkularitas.
BBL	<i>Borough-Block-Lot</i> , kode identifikasi unik lot pajak di Kota New York yang menjadi kunci penggabungan antar-dataset.
PLUTO	<i>Primary Land Use Tax Lot Output</i> , basis data properti yang memuat luas, jumlah unit, tahun bangun, dan kepemilikan tiap lot.
HPD	<i>Housing Preservation and Development</i> , badan kota yang melakukan inspeksi dan mencatat pelanggaran kode perumahan.
ACS	<i>American Community Survey</i> , survei sensus AS yang menyediakan indikator sosio-ekonomi tingkat wilayah.
ZCTA	<i>ZIP Code Tabulation Area</i> , satuan wilayah sensus yang mendekati batas kode pos untuk penggabungan dengan ACS.

(lanjutan Glosarium)

Istilah	Definisi
HDFC	<i>Housing Development Fund Corporation</i> , badan nirlaba pengelola stok perumahan terjangkau.
Deviasi komposisi	Selisih antara komposisi pengaduan suatu wilayah dengan komposisi rata-rata seluruh kota pada tahun yang sama.
Bias pelaporan	Kecenderungan perilaku melapor warga yang berbeda-beda antarwilayah sehingga data pengaduan tidak setara dengan kondisi objektif.
Sirkularitas administratif	Keadaan saat pengaduan 311 memicu inspeksi HPD, sehingga pelanggaran tercatat sebagian merupakan respons atas pengaduan itu sendiri.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Setiap hari, warga Kota New York mengirimkan ratusan ribu pengaduan melalui layanan telepon dan aplikasi *311* mengenai berbagai masalah di lingkungan mereka, mulai dari pemanas ruangan yang mati, pipa bocor, sampah yang menumpuk, hingga kendaraan yang diparkir sembarangan. Selama bertahun-tahun, jutaan pengaduan ini terkumpul menjadi sebuah jejak digital yang sangat besar tentang denyut keseharian kota. Data semacam ini telah dimanfaatkan untuk berbagai keperluan, mulai dari pemantauan kondisi lingkungan hingga penilaian kualitas layanan kota (Li dan Hyder, 2020). Pertanyaan yang menarik muncul dari sini, yaitu apakah tumpukan pengaduan tersebut hanya catatan administratif biasa, ataukah di dalamnya tersimpan pola yang bermakna tentang kondisi nyata bangunan dan wilayah.

Sebuah bangunan yang terabaikan biasanya tidak rusak pada satu sistem saja. Ketika pemilik berhenti merawat sebuah gedung, masalah cenderung muncul beruntun, misalnya pemanas yang bermasalah sering berdampingan dengan pipa yang bocor, cat yang mengelupas, dan kondisi yang tidak higienis. Bila pola beruntun semacam itu benar-benar ada, maka kemunculan bersama berbagai jenis pengaduan pada satu bangunan dapat berfungsi sebagai semacam sidik jari pengabaian. Penelitian terdahulu menunjukkan bahwa pengaduan warga memang dapat menjadi penanda kerusakan fisik dan tekanan sosial di tingkat lingkungan (Li dan Hyder, 2020; Wang dkk., 2017).

Data *311* dipilih karena tiga alasan. Pertama, ukurannya sangat besar, yaitu lebih dari dua puluh juta catatan, sehingga cocok untuk teknik penambangan data (Pazdora dkk., 2022). Kedua, data ini dapat dihubungkan dengan sumber lain yang menggambarkan karakteristik fisik bangunan dan catatan pelanggaran resmi, sehingga pola yang ditemukan dapat divalidasi. Ketiga, dan yang paling penting secara metodologis, data *311* hanya merekam apa yang dilaporkan warga, bukan apa yang sebenarnya terjadi. Perilaku melapor warga diketahui berbeda-beda antarwilayah dan dapat dipengaruhi oleh faktor sosial-ekonomi, sehingga data pengaduan mengandung bias pelaporan yang harus disikapi secara hati-hati (Kontokosta dan Hong, 2021). Perbedaan antara yang dilaporkan dan yang sebenarnya terjadi inilah yang menjadi lensa utama yang menyertai seluruh penafsiran dalam laporan ini.

1.2 Analisis Kesenjangan Literatur

Telaah terhadap penelitian terdahulu memperlihatkan tiga kesenjangan yang belum terjawab. Kesenjangan pertama adalah skala analisis. Li dan Hyder (2020) maupun Wang dkk. (2017) bekerja pada tingkat wilayah, yaitu mengagregasi pengaduan per lingkungan atau per kode pos lalu menghubungkannya dengan karakteristik sosio-ekonomi. Pendekatan ini berguna untuk memetakan tekanan kawasan, tetapi tidak menjawab apakah pada satu bangunan tertentu jenis-jenis pengaduan saling menempel membentuk pola yang bermakna. Tingkat bangunan individual sebagai unit analisis penambangan asosiasi belum tergarap.

Kesenjangan kedua adalah desain validasi. Banyak penelitian memanfaatkan data pengaduan sebagai proksi kondisi, namun Kontokosta dan Hong (2021) menegaskan bahwa data pengaduan sarat bias pelaporan. Persoalan menjadi lebih rumit karena pengaduan perumahan diteruskan ke badan inspeksi, sehingga catatan pelanggaran resmi sebagian merupakan akibat dari pengaduan itu sendiri. Belum ada penelitian yang secara eksplisit merancang validasi untuk memisahkan sinyal nyata dari sirkularitas administratif ini.

Kesenjangan ketiga adalah dimensi temporal. Penelitian terdahulu jarang menganalisis urutan kemunculan jenis pengaduan, yaitu jenis pengaduan mana yang cenderung dilaporkan lebih dulu dan mana yang menyusul. Padahal urutan ini berpotensi mengungkap dinamika pengabaian. Ketiga kesenjangan inilah yang menjadi ruang kontribusi penelitian ini.

1.3 Rumusan Masalah

Penelitian ini berangkat dari dua pertanyaan penelitian (*research question*, RQ) yang bersifat terbuka. Sesuai prinsip penemuan dan bukan konfirmasi yang memandu seluruh penelitian, kedua pertanyaan dirumuskan sebagai pertanyaan murni tanpa menetapkan arah jawaban di muka. Arah jawaban dibiarkan muncul dari data, dan setiap pola yang ditemukan diuji terhadap kemungkinan penjelasan lain sebelum diterima.

RQ1. Pada satu bangunan yang sama, apakah jenis-jenis pengaduan tertentu muncul bersama melebihi yang diharapkan secara kebetulan, sehingga membentuk pola yang dapat ditafsirkan sebagai sidik jari pengabaian bangunan?

RQ2. Apakah wilayah-wilayah di Kota New York berbeda dalam komposisi jenis pengaduannya, dan apakah komposisi itu bergeser dari waktu ke waktu?

1.4 Kriteria Keberhasilan

Agar penafsiran temuan tidak bias, kriteria yang menentukan apakah sebuah temuan dianggap bermakna ditetapkan sebelum analisis dijalankan, sebagaimana dirangkum pada Tabel 1.

Tabel 1 Kriteria keberhasilan yang ditetapkan sebelum analisis

Pertanyaan	Kriteria temuan dianggap valid
RQ1	Nilai <i>lift</i> antarpasangan item inti lebih dari satu pada seluruh data, dan asosiasi dengan pelanggaran HPD bertahan (tetap meningkat seiring lengkapan sindrom) pada jendela prior yang independen.
RQ2	Rata-rata <i>silhouette</i> pengelompokan lebih dari 0,15, dan komposisi antar cluster berbeda secara substansial pada minimal tiga jenis pengaduan teratas tiap cluster.

Sumber: penetapan peneliti sebelum analisis

1.5 Tujuan Penelitian

Sesuai dengan rumusan masalah, penelitian ini memiliki dua tujuan. Pertama, menemukan dan memvalidasi pola kemunculan bersama jenis pengaduan pada tingkat bangunan. Kedua, menemukan dan mengkarakterisasi tipologi wilayah berdasarkan komposisi pengaduannya beserta pergeserannya dari waktu ke waktu.

1.6 Manfaat Penelitian

Penelitian ini berkontribusi pada pemahaman bahwa data pengaduan warga dapat berfungsi sebagai instrumen diagnostik kondisi bangunan. Secara praktis, temuan dapat menjadi dasar bagi penentuan prioritas inspeksi perumahan. Secara metodologis, penelitian ini menunjukkan bagaimana sebuah analisis dapat dirancang agar mampu menguji dan membuktikan dirinya sendiri salah, sebuah sikap yang menjaga kejujuran ilmiah.

1.7 Ruang Lingkup

Penelitian ini dibatasi pada metode penambangan data deskriptif, yaitu penambangan aturan asosiasi (*association rule mining*) dan pengelompokan (*clustering*). Metode yang bersifat prediktif seperti regresi dan *forecasting* berada di luar lingkup, karena tergolong *machine learning*. Konsekuensi penting dari batasan ini adalah bahwa setiap variabel pengganggu (*confounding*) dikendalikan melalui stratifikasi, bukan

melalui pemodelan regresi. Objek penelitian adalah data *311* Kota New York periode 2020–2024, yang diperkaya dengan data properti *PLUTO*, data pelanggaran perumahan *HPD*, dan data sensus *ACS*.

BAB II

DATA PROVENANCE DAN AUDIT

2.1 Asal-Usul Data dan Rantai Pengolahan

Data *NYC 311 Service Requests* dikumpulkan oleh Pemerintah Kota New York melalui kanal layanan non-darurat 311 yang mencakup telepon, aplikasi seluler, dan portal web. Motivasi pengumpulannya adalah transparansi pemerintahan dan responsivitas layanan, bukan penelitian, sehingga struktur datanya mengikuti alur kerja administratif. Data diperbarui harian dan dipublikasikan melalui portal NYC Open Data; versi yang digunakan dalam penelitian ini mencakup periode 2020 hingga 2024. Bias sistemik yang melekat adalah bias pelaporan, yaitu data hanya memuat masalah yang cukup mengganggu warga untuk dilaporkan dan yang warganya cukup percaya pada layanan kota untuk melapor (Kontokosta dan Hong, 2021).

Data *PLUTO* dikumpulkan oleh Department of City Planning untuk keperluan perencanaan tata guna lahan. Data ini memuat atribut fisik tiap lot pajak, seperti luas bangunan, jumlah unit hunian, tahun pembangunan, dan nama pemilik. Pembaruannya berkala beberapa kali setahun. Bias sistemiknya bersumber pada kesalahan administratif penulisan, misalnya format angka dan artefak teks pada nama pemilik, yang ditangani pada tahap pembersihan.

Data *HPD* dikumpulkan oleh Department of Housing Preservation and Development dari hasil inspeksi pelanggaran kode perumahan. Setiap catatan pelanggaran memiliki kelas keparahan. Inilah sumber yang berfungsi sebagai validator independen, karena pelanggaran dicatat oleh inspektur, bukan oleh warga. Namun justru pada relasi antara 311 dan HPD muncul persoalan sirkularitas. Pengaduan perumahan melalui 311 diteruskan ke HPD untuk diinspeksi, dan inspeksi itulah yang menghasilkan catatan pelanggaran. Akibatnya, pelanggaran HPD yang tercatat sezaman dengan pengaduan sebagian merupakan respons administratif terhadap pengaduan itu sendiri. Kondisi inilah yang menjadi dasar perancangan desain dua jendela waktu pada metodologi.

Data *ACS* dikumpulkan oleh U.S. Census Bureau sebagai survei berkelanjutan atas kondisi sosio-ekonomi penduduk. Estimasi lima tahunan digunakan karena lebih stabil pada satuan wilayah kecil. Data ini diagregasi pada satuan ZCTA yang mendekati batas kode pos. Bias sistemiknya adalah ketidaktepatan padanan antara batas ZCTA dan batas kode pos administratif, serta sifat estimasi survei yang mengandung galat sampling.

2.2 Kamus Data

Tabel 2 menyajikan kamus data untuk variabel kunci *NYC 311* beserta catatan kritisnya. Tabel 3 merangkum variabel kunci dan limitasi tiga sumber pendukung.

Tabel 2 Kamus data variabel kunci NYC 311 Service Requests

Variabel	Tipe	Deskripsi dan catatan kritis	Hilang
UniqueKey	Integer	Pengenal unik tiap pengaduan; diperiksa dan tidak ditemukan duplikat.	0%
CreatedDate	Datetime	Tanggal pengaduan dibuat; dasar seluruh analisis temporal.	0%
ClosedDate	Datetime	Tanggal pengaduan ditutup; hilang 28,3% karena berstatus <i>Open</i> , bersifat struktural bukan acak.	28,3%
ComplaintType	Kategori	Jenis pengaduan; basis instrumen sindrom dan tipologi wilayah.	0%
Descriptor	Kategori	Rincian sub-jenis pengaduan.	1,2%
LocationType	Kategori	Tipe lokasi kejadian (mis. bangunan, jalan).	6,4%
IncidentZip	Kategori	Kode pos kejadian; basis agregasi wilayah RQ2.	3,0%
BBL	Kategori	Kode lot bangunan; kunci unit analisis. Hilang 11,66% dengan mekanisme MNAR (lihat 2.4).	11,66%
Borough	Kategori	Wilayah administratif kota.	0%
Status	Kategori	Status penyelesaian pengaduan.	0%
Latitude	Float	Lintang lokasi; hilang 5,1% sebagian akibat penyamaran privasi.	5,1%
Longitude	Float	Bujur lokasi; pola hilang sama dengan Latitude.	5,1%

Sumber: NYC Open Data (2024), diolah

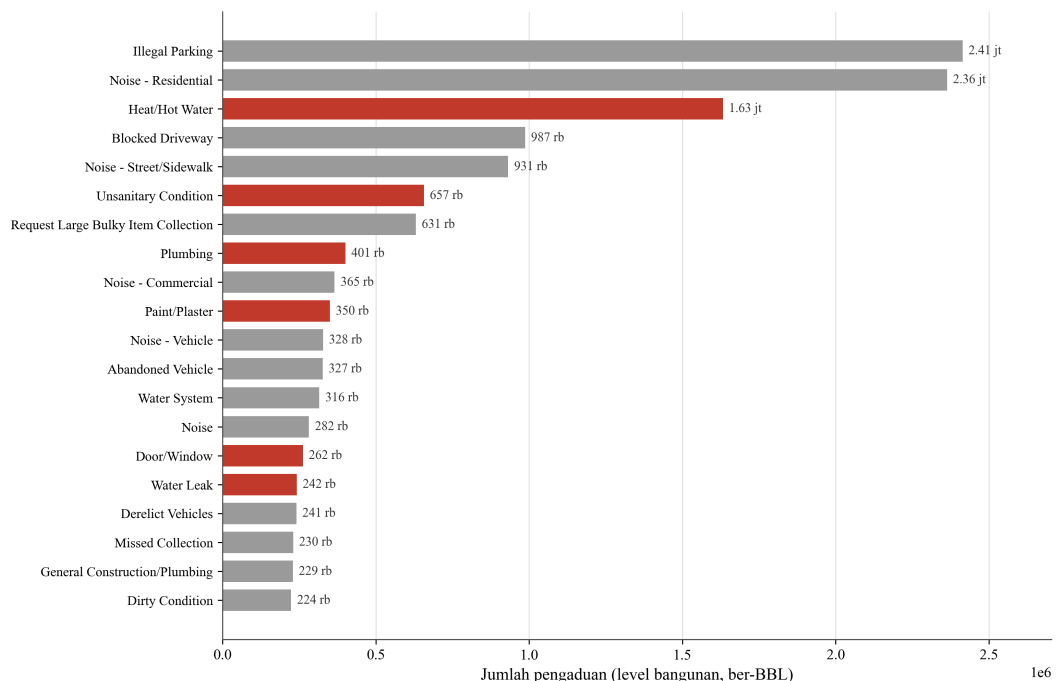
Tabel 3 Variabel kunci dan limitasi sumber data pendukung

Sumber	Variabel kunci	Catatan limitasi
PLUTO	BBL, UnitsRes, BldgArea, YearBuilt, OwnerName	LotArea/BldgArea sempat salah baca akibat pemisah ribuan; OwnerName mengandung artefak spasi.
HPD	BBL, Class, InspectionDate	Hanya memuat pelanggaran yang sempat diinspeksi; tunduk pada sirkularitas dengan 311.
ACS	ZCTA, Median Income, Rent, % Sarjana, % Penyewa	Estimasi survei dengan galat sampling; batas ZCTA tidak persis sama dengan kode pos.

Sumber: NYC Open Data dan U.S. Census Bureau (2024), diolah

2.3 Analisis Eksploratori Data

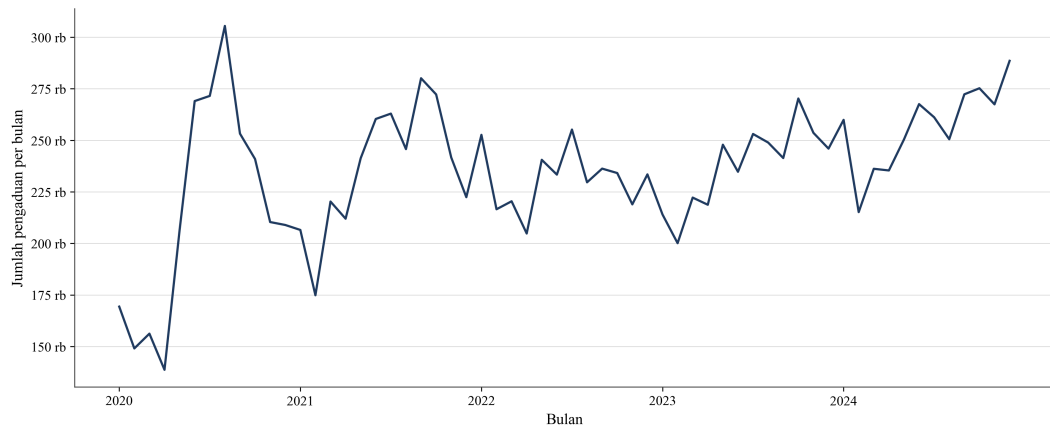
Sebelum penambangan, dilakukan analisis eksploratori untuk memahami bentuk data. Gambar 1 menampilkan distribusi dua puluh jenis pengaduan teratas. Distribusi ini berekor panjang, yaitu sejumlah kecil jenis pengaduan mendominasi volume sementara sebagian besar jenis lain jarang muncul. Enam dari dua puluh jenis teratas inilah yang kelak membentuk instrumen sindrom.



Gambar 1 Distribusi dua puluh jenis pengaduan teratas berdasarkan volume

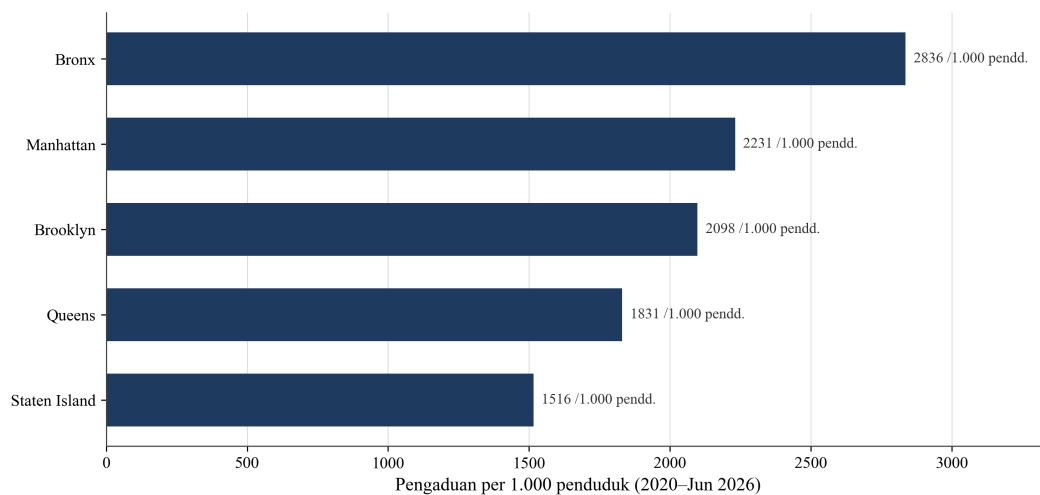
Gambar 2 menampilkan distribusi temporal pengaduan bulanan sepanjang 2020 hingga 2024. Pada tahun 2020 terlihat anomali yang berkaitan dengan pandemi, yang

memengaruhi pola pelaporan dan menjadi alasan kehati-hatian saat menafsirkan tren awal periode.

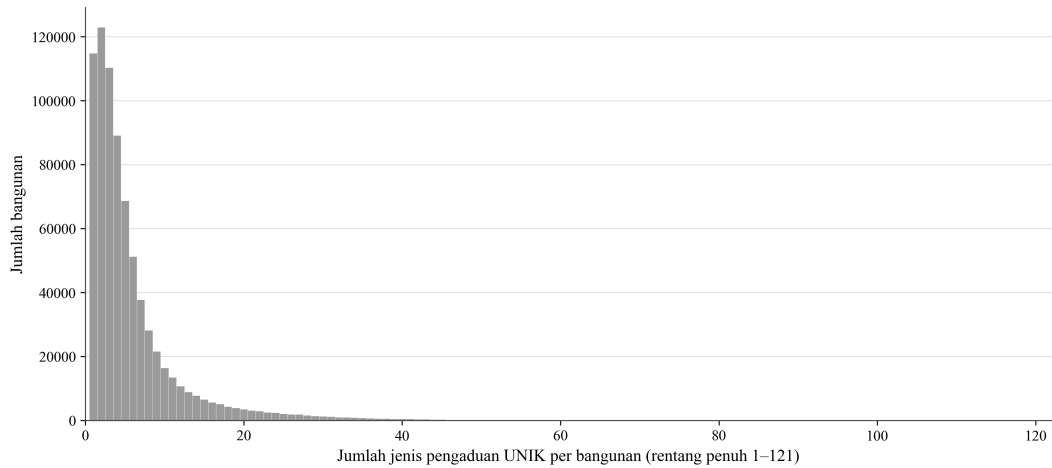


Gambar 2 Distribusi temporal volume pengaduan bulanan periode 2020–2024

Gambar 3 menampilkan distribusi geografis pengaduan per borough dalam bentuk peta densitas, yang memperlihatkan konsentrasi pelaporan yang tidak merata antarwilayah. Gambar 4 menampilkan distribusi jumlah jenis pengaduan per keranjang bangunan sebelum penyaringan sindrom, yang menjadi pembanding bagi distribusi pasca-sindrom pada Bab V.



Gambar 3 Distribusi geografis densitas pengaduan antar-borough



Gambar 4 Distribusi jumlah jenis pengaduan per keranjang bangunan sebelum penyaringan sindrom

2.4 Analisis Mekanisme Data Hilang

Variabel BBL memerlukan analisis mendalam karena nilainya menentukan apakah sebuah catatan dapat masuk ke unit analisis bangunan. Tiga mekanisme data hilang ditinjau. Mekanisme MCAR, yaitu hilang sepenuhnya acak, secara konseptual tidak berlaku di sini. Walaupun uji Little tersedia untuk menguji MCAR secara formal, sudah jelas bahwa hilangnya BBL berkorelasi dengan *ComplaintType*, sehingga asumsi acak penuh gugur. Mekanisme MAR, yaitu hilang bergantung variabel terobservasi lain, berlaku sebagian, karena pengaduan bertipe jalanan seperti *Noise - Street/Sidewalk* dan *Illegal Parking* secara struktural memang tidak memiliki BBL. Mekanisme MNAR, yaitu hilang bergantung nilai variabel itu sendiri, berlaku untuk bangunan dengan alamat tidak terdaftar, yang justru berpotensi merupakan bangunan bermasalah.

Keputusan penanganan ditetapkan berdasarkan peran variabel. Karena BBL adalah kunci identifikasi unit analisis dan bukan variabel prediktor, imputasi tidak tepat sebab tidak ada makna mengimputasi identitas bangunan. Keputusan yang diambil adalah penghapusan menyeluruh baris tanpa BBL (*listwise deletion*), yang menghapus 11,66 persen catatan dan menghasilkan 18,85 juta catatan terpakai. Untuk Latitude dan Longitude yang hilang 5,1 persen, sebagian besar terkait penyamaran privasi tipe lokasi tertentu; karena keduanya tidak menjadi kunci penggabungan utama, ketidaklengkapan ini tidak memengaruhi unit analisis bangunan dan hanya membatasi analisis spasial titik.

2.5 Pernyataan Keterwakilan Data

Keterwakilan data dinyatakan secara eksplisit dalam dua arah. Dalam arah positif, dataset ini merepresentasikan pengaduan yang dilaporkan warga Kota New York melalui kanal 311, baik telepon, aplikasi, maupun portal web, selama periode 2020 hingga 2024, yang berhasil dipetakan ke lot bangunan teridentifikasi.

Dalam arah negatif, dataset ini tidak merepresentasikan beberapa hal yang penting untuk diakui. Pertama, kondisi fisik bangunan yang terabaikan namun tidak pernah dilaporkan, yang dapat disebut materi gelap pengabaian (*dark matter of neglect*). Kedua, masalah yang diselesaikan secara informal sebelum masuk sistem. Ketiga, komunitas dengan tingkat kepercayaan rendah terhadap layanan pemerintah, yang cenderung melapor lebih sedikit sehingga terwakili secara kurang. Keempat, stok bangunan informal tanpa BBL terdaftar. Ketidakwakilan ini bukan kelemahan yang dapat diperbaiki dengan teknik analisis, melainkan batas ontologis data 311 yang harus menyertai seluruh penafsiran.

BAB III

TINJAUAN PUSTAKA

3.1 Data 311 sebagai Penanda Kondisi Perkotaan

Layanan 311 merupakan kanal pengaduan non-darurat yang disediakan banyak kota untuk warganya. Karena bersifat terbuka dan diperbarui secara berkala, data ini banyak dimanfaatkan dalam penelitian perkotaan. Li dan Hyder (2020) menunjukkan bahwa jenis-jenis pengaduan 311 dapat berfungsi sebagai indikator tekanan lingkungan, baik berupa kekacauan sosial maupun kerusakan fisik. Pada tingkat wilayah, Wang dkk. (2017) memperlihatkan bahwa komposisi jenis pengaduan dapat menjadi tanda tangan khas sebuah lokasi yang berhubungan dengan karakteristik sosio-ekonominya. Penelitian-penelitian ini menjadi dasar gagasan bahwa pengaduan warga memuat sinyal nyata mengenai kondisi kota.

Akan tetapi, data 311 mengandung keterbatasan mendasar. Kontokosta dan Hong (2021) menemukan bahwa perilaku melapor warga berbeda-beda secara sistematis antarwilayah, sehingga data pengaduan tidak dapat langsung diartikan sebagai ukuran objektif atas kondisi yang sebenarnya. Temuan ini menegaskan perlunya validasi pola pengaduan terhadap sumber data yang independen, sebuah prinsip yang dijadikan landasan utama dalam penelitian ini.

3.2 Penambangan Aturan Asosiasi

Penambangan aturan asosiasi adalah teknik untuk menemukan keterkaitan yang sering muncul bersama dalam sebuah kumpulan transaksi. Teknik ini awalnya populer untuk analisis keranjang belanja, yaitu menemukan barang-barang yang cenderung dibeli bersamaan. Dalam penelitian ini, keranjang bukanlah belanjaan, melainkan sebuah bangunan, dan barang di dalamnya adalah jenis-jenis pengaduan yang pernah dilaporkan untuk bangunan tersebut.

Tiga ukuran utama digunakan untuk menilai sebuah aturan asosiasi (Hashad dkk., 2024). *Support* mengukur seberapa sering sebuah kombinasi muncul di seluruh data. *Confidence* mengukur seberapa sering kombinasi muncul jika salah satu anggotanya sudah ada. *Lift* mengukur seberapa besar dua hal muncul bersama melebihi yang diharapkan secara kebetulan. Nilai *support* untuk kombinasi item A dan B dirumuskan sebagai berikut.

$$\text{support}(A \cup B) = \frac{n(A \cup B)}{N} \quad (1)$$

dengan,

$n(A \cup B)$ = jumlah keranjang yang memuat item A dan B sekaligus,

N = jumlah seluruh keranjang.

Selanjutnya, nilai *lift* dihitung sebagai perbandingan antara *support* gabungan terhadap perkalian *support* masing-masing item.

$$\text{lift}(A, B) = \frac{\text{support}(A \cup B)}{\text{support}(A) \cdot \text{support}(B)} \quad (2)$$

dengan,

$\text{support}(A)$ = proporsi keranjang yang memuat item A ,

$\text{support}(B)$ = proporsi keranjang yang memuat item B .

Nilai *lift* sama dengan satu menandakan dua item muncul bersama persis seperti yang diharapkan jika keduanya saling bebas. Nilai *lift* lebih dari satu menandakan keterkaitan yang nyata, yaitu kedua item muncul bersama lebih sering daripada kebetulan. Dalam penelitian ini, *lift* dipilih sebagai gerbang utama karena ia paling tepat menerjemahkan gagasan melebihi kebetulan, sedangkan *confidence* dapat menyesatkan ketika salah satu jenis pengaduan memang sangat umum. Algoritma yang digunakan untuk menemukan kombinasi yang sering muncul adalah *FP-Growth*, sebuah metode yang efisien karena tidak perlu memeriksa seluruh kemungkinan kombinasi satu per satu (Hashad dkk., 2024).

3.3 Pengelompokan dengan K-Means

Pengelompokan adalah teknik untuk mengelompokkan objek-objek yang mirip ke dalam kelompok yang sama, tanpa label yang ditentukan sebelumnya. Algoritma *K-means* membagi objek ke dalam sejumlah k kelompok dengan cara meminimalkan jarak setiap objek ke pusat kelompoknya. Kualitas pengelompokan dapat dinilai dengan nilai *silhouette*, yang untuk satu objek i dirumuskan sebagai berikut.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

dengan,

$a(i)$ = rata-rata jarak objek i ke seluruh objek lain dalam kelompok yang sama,

$b(i)$ = rata-rata jarak terkecil objek i ke objek-objek pada kelompok tetangga terdekat.

Nilai *silhouette* berkisar antara -1 dan 1 . Nilai mendekati satu menandakan kelompok yang terpisah jelas, sedangkan nilai rendah menandakan kelompok yang

saling bergradasi. Ukuran ini menjadi dasar penilaian apakah tipologi wilayah yang ditemukan bersifat tegas atau berupa peralihan yang halus.

3.4 Pendekatan CRISP-DM

Penelitian ini mengikuti kerangka *Cross-Industry Standard Process for Data Mining* (CRISP-DM), yang membagi proses menjadi tahapan pemahaman masalah, pemahaman data, persiapan data, pemodelan, evaluasi, dan komunikasi. Kerangka ini memastikan analisis berjalan sistematis dan dapat ditelusuri kembali.

3.5 Prinsip Penemuan, Bukan Konfirmasi

Prinsip yang memandu seluruh penelitian adalah penemuan, bukan konfirmasi. Artinya, ketidaktahuan berlaku untuk jawaban, bukan untuk metode. Memilih objek dan metode analisis adalah keputusan lingkup yang sah, tetapi menetapkan arah jawaban sebelum data berbicara adalah hal yang dihindari. Selain itu, angka-angka deskriptif diperlakukan sebagai pemeriksaan kelayakan, bukan sebagai temuan akhir. Data eksternal sengaja dimasukkan hanya setelah pola ditemukan dari data *311* murni, untuk mencegah bias. Pendekatan ini sejalan dengan kesadaran bahwa data pengaduan mengandung bias pelaporan, sehingga setiap pola wajib diuji terhadap kemungkinan penjelasan lain (Kontokosta dan Hong, 2021).

BAB IV

METODOLOGI PENELITIAN

4.1 Jenis dan Sumber Data

Penelitian ini menggunakan empat sumber data dengan peran yang berbeda, sebagaimana dirangkum pada Tabel 4. Pembagian peran ini penting, yaitu data *311* digunakan untuk penemuan pola, data *PLUTO* untuk pengayaan dan pengendalian variabel pengganggu, data *HPD* untuk validasi independen, dan data *ACS* untuk pemeriksaan konteks sosio-ekonomi.

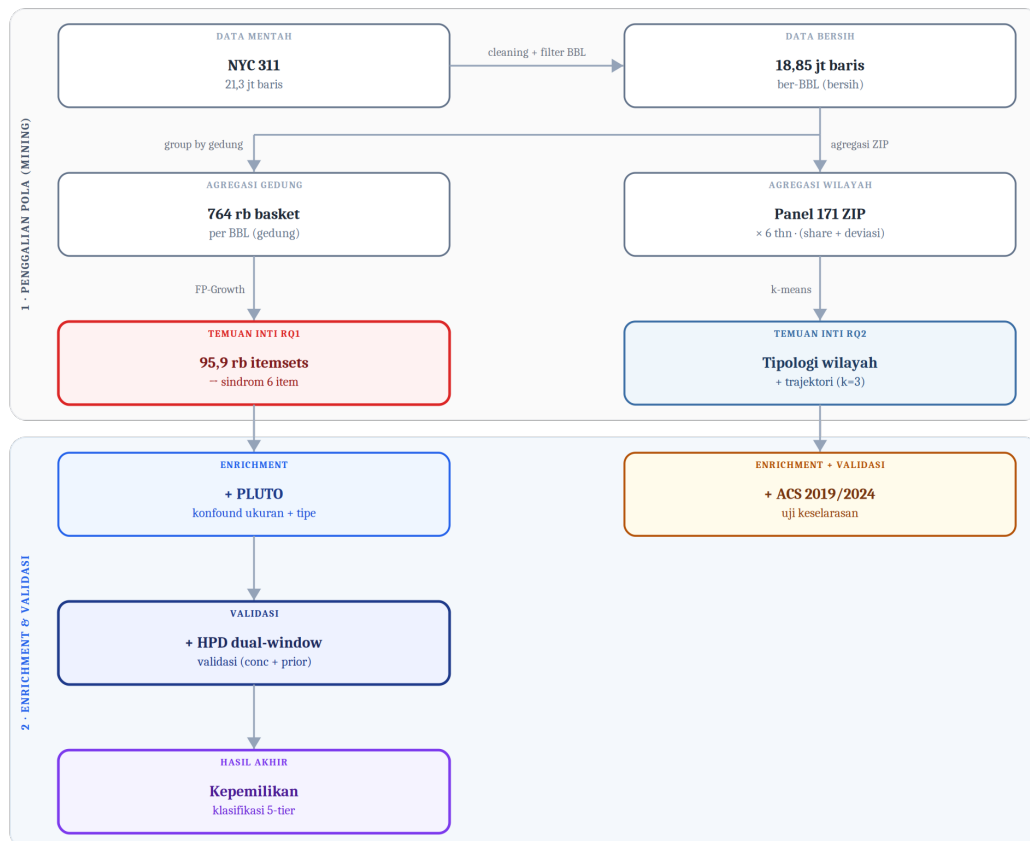
Tabel 4 Empat sumber data dan perannya dalam penelitian

Sumber	Volume	Peran
NYC 311	21,3 juta catatan	Penemuan pola (penambangan)
PLUTO	858.602 lot	Pengayaan dan pengendalian ukuran serta tipe bangunan
HPD	11,0 juta pelanggaran	Validasi independen oleh inspektur
ACS	2 periode per ZCTA	Konteks perubahan sosio-ekonomi

Sumber: NYC Open Data dan U.S. Census Bureau (2024), diolah

4.2 Alur Pengolahan Data

Alur kerja secara keseluruhan ditampilkan pada Gambar 5. Data *311* mentah dibersihkan dan disaring hingga tersisa catatan yang dapat dikaitkan dengan kode lot bangunan (BBL). Catatan ini kemudian diubah menjadi format keranjang, yaitu satu baris untuk setiap bangunan berisi daftar jenis pengaduan unik yang pernah dilaporkan. Dari format inilah pola asosiasi ditambang. Setelah pola ditemukan, data eksternal dimasukkan secara bertahap untuk pengayaan dan validasi. Penataan ini secara sengaja memisahkan fase eksploratif yang murni dari data *311* dengan fase konfirmatori yang melibatkan data eksternal.



Gambar 5 Alur pengolahan data dari 21,3 juta pengaduan hingga temuan tervalidasi

4.3 Persiapan dan Pembersihan Data

Tahap pembersihan menghasilkan beberapa temuan kualitas data yang penting untuk didokumentasikan, karena masing-masing berpotensi merusak hasil bila tidak ditangani. Pertama, kolom luas bangunan pada data *PLUTO* ternyata menggunakan tanda titik sebagai pemisah ribuan, sehingga nilai seperti 1.683 yang sebenarnya berarti 1.683 kaki persegi terbaca sekitar seribu kali lebih kecil. Kedua, terdapat lima belas pasang jenis pengaduan yang sama tetapi berbeda penulisan huruf besar-kecil, yang menciptakan kemunculan-bersama palsu. Ketiga, banyak nama pemilik perumahan terjangkau pada *PLUTO* mengandung spasi yang tersisip akibat kesalahan penulisan sumber. Ketiga masalah ini ditemukan karena diterapkan prosedur menampilkan nilai mentah sebelum konversi tipe data. Rincian kronologis seluruh keputusan pembersihan disajikan pada Lampiran B.

Penyaringan berdasarkan kode bangunan menghapus sekitar 11,66 persen catatan yang tidak dapat dipetakan ke lot tertentu, yang umumnya berupa pengaduan jalanan yang tidak terkait properti. Setelah pembersihan, tersisa 18,85 juta catatan yang membentuk 764.114 keranjang bangunan.

4.4 Justifikasi Pemilihan Algoritma

Pemilihan algoritma penambangan asosiasi mempertimbangkan tiga kandidat utama, yaitu Apriori, ECLAT, dan FP-Growth, dengan perbandingan pada Tabel 5. FP-Growth dipilih karena tidak memerlukan pembangkitan kandidat kombinasi secara eksplisit, menggunakan struktur pohon yang efisien, dan terbukti berskala baik pada dataset besar. Trade-off yang diterima adalah bahwa FP-tree dapat menuntut memori lebih besar dibanding ECLAT pada data yang sangat padat. Namun karena data ini bersifat *sparse*, yaitu jumlah jenis pengaduan per bangunan relatif sedikit dibanding total jenis pengaduan yang ada, FP-Growth tetap lebih efisien.

Tabel 5 Perbandingan tiga algoritma kandidat penambangan asosiasi

Aspek	Apriori	ECLAT	FP-Growth
Kompleksitas waktu	Tinggi, banyak pemindaian basis data	Sedang	Rendah, dua pemindaian saja
Kebutuhan memori	Rendah	Rendah	Sedang hingga tinggi (FP-tree)
Skalabilitas > 18 juta catatan	Buruk	Cukup	Baik
Transparansi proses	Tinggi	Sedang	Sedang
Kemudahan interpretasi	Tinggi	Sedang	Tinggi

Sumber: sintesis peneliti berdasarkan Hashad dkk. (2024)

4.5 Instrumen Sindrom Pemeliharaan Interior

Dari hasil penambangan, enam jenis pengaduan secara konsisten mendominasi kombinasi-kombinasi yang paling sering muncul bersama (yaitu yang ber-*support* tinggi), yaitu Pemanas/Air Panas (*Heat/Hot Water*), Kondisi Tidak Higienis (*Unsanitary Condition*), Pipa (*Plumbing*), Cat/Plester (*Paint/Plaster*), Kebocoran Air (*Water Leak*), dan Pintu/Jendela (*Door/Window*). Perlu ditegaskan bahwa keenam item ini dipilih atas dasar frekuensi yang tinggi dan relevansi domain terhadap pemeliharaan bangunan, bukan atas dasar nilai *lift* tertinggi; aturan ber-*lift* tertinggi pada data justru berasal dari ranah layanan makanan komersial yang tidak relevan dengan pertanyaan penelitian ini. Keenam jenis ini diangkat menjadi instrumen pengukuran yang disebut sindrom pemeliharaan interior. Pemilihan ini sejalan dengan temuan bahwa pengaduan terkait perawatan bangunan merupakan penanda kerusakan fisik

lingkungan (Li dan Hyder, 2020).

Instrumen utama berupa variabel hitungan yang bernilai nol sampai enam, yaitu berapa banyak dari keenam jenis pengaduan tersebut yang pernah dilaporkan untuk sebuah bangunan. Variabel kontinu ini dipilih sebagai instrumen utama karena memungkinkan analisis hubungan dosis-respons yang tidak bergantung pada satu ambang tertentu. Sebagai label komunikasi sekunder, digunakan penanda biner pada ambang empat dari enam. Ambang yang inklusif ini dipilih secara sengaja, karena data *311* hanya merekam pengaduan yang dilaporkan, sehingga ambang yang terlalu ketat berisiko mengabaikan bangunan yang sebenarnya bermasalah namun pelaporannya tidak lengkap.

4.6 Spesifikasi Parameter dan Analisis Sensitivitas

Parameter penambangan ditetapkan secara eksplisit beserta justifikasinya. Parameter *min_support* ditetapkan sebesar 0,005 setelah mengevaluasi tiga nilai kandidat, yaitu 0,001 (rendah), 0,005 (sedang), dan 0,01 (tinggi). Pada *min_support* 0,001, jumlah kombinasi membengkak menjadi 2.210.478 sehingga enumerasi aturannya menjadi tidak praktis secara komputasi dan dihentikan; pada 0,01, jumlah kombinasi turun menjadi 16.454 sehingga sebagian pola yang masih relevan secara domain hilang. Nilai 0,005, yang menghasilkan 95.902 kombinasi, dipilih sebagai keseimbangan antara cakupan pola dan ketertelusuran. Parameter *min_confidence* ditetapkan 0,5 dan *min_lift* substantif ditetapkan 1,5 (lihat 4.9). Tabel 6 merangkum analisis sensitivitas terhadap nilai-nilai kandidat tersebut.

Tabel 6 Analisis sensitivitas parameter *min_support* penambangan asosiasi

min_support	Jumlah kombinasi	Aturan lift > 1,5	Keterangan
0,001 (rendah)	2.210.478	tidak dihitung	Enumerasi aturan tak praktis (dihentikan)
0,005 (terpilih)	95.902	2.132.126	Titik keseimbangan
0,010 (tinggi)	16.454	309.920	Kehilangan sebagian kombinasi relevan

Sumber: hasil penambangan (seed 42); enumerasi aturan pada 0,001 dihentikan karena tidak praktis

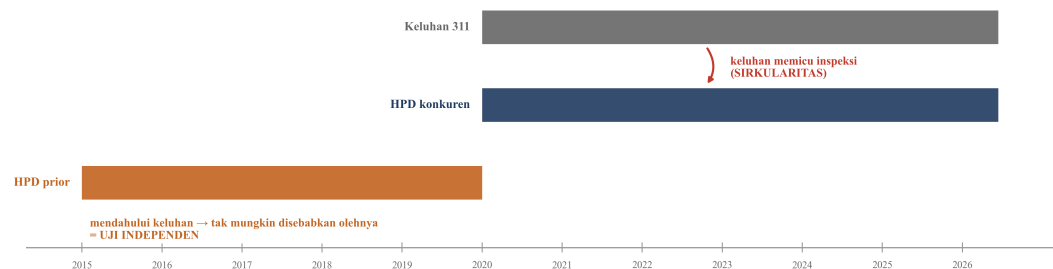
Untuk pengelompokan *K-means*, jumlah kelompok ditetapkan $k = 3$. Penting untuk dinyatakan secara jujur bahwa $k = 2$ justru memberikan nilai *silhouette* tertinggi (0,27 berbanding 0,22 pada $k = 3$). Namun $k = 2$ menggabungkan tipe-tipe wilayah yang secara domain berbeda menjadi satu, sehingga $k = 3$ dipilih karena

menghasilkan tipologi yang lebih kaya dan dapat diinterpretasikan, bukan karena memaksimalkan *silhouette*. Nilai *silhouette* pada $k = 3$ tetap melampaui ambang keberhasilan 0,15 yang ditetapkan di muka. Kestabilan diuji pada 20 nilai *random seed*, dan variasi *silhouette* antar-*seed* sangat kecil (standar deviasi 0,0001). Rincian perbandingan untuk $k = 2$ sampai $k = 5$ disajikan pada Lampiran C.

4.7 Desain Validasi Dua Jendela

Tantangan utama validasi adalah kemungkinan sirkularitas. Di Kota New York, pengaduan perumahan melalui *311* diteruskan ke badan HPD untuk diinspeksi, dan inspeksi itulah yang menghasilkan catatan pelanggaran. Akibatnya, pelanggaran HPD pada periode yang sama dengan pengaduan sebagian disebabkan oleh pengaduan itu sendiri. Bila validasi hanya menggunakan periode yang bersamaan, hubungan yang ditemukan menjadi sebagian sirkular.

Untuk mengatasi hal ini, validasi menggunakan dua jendela waktu, sebagaimana digambarkan pada Gambar 6. Jendela konkuren mencakup pelanggaran sejak 2020 yang sezaman dengan pengaduan. Jendela prior mencakup pelanggaran 2015 sampai 2019, yang mendahului pengaduan sehingga tidak mungkin disebabkan olehnya. Bila pola tetap muncul pada jendela prior, maka pola itu konsisten dengan pengabaian kronis yang nyata, bukan sekadar gema administratif. Desain ini merupakan penerapan langsung dari prinsip bahwa pola pada data pengaduan harus diuji terhadap sumber independen (Kontokosta dan Hong, 2021).



Gambar 6 Desain dua jendela untuk memisahkan sinyal nyata dari sirkularitas administratif

4.8 Definisi Baseline

Agar temuan dapat ditafsirkan, ditetapkan dua baseline. Baseline teoritis untuk asosiasi adalah nilai *lift* sama dengan satu, yang secara definisional menunjukkan dua item saling bebas. Setiap aturan dengan *lift* lebih dari satu melebihi kebetulan, namun agar dianggap substantif, ditetapkan gerbang *lift* lebih dari 1,5 berdasarkan praktik umum

dalam literatur. Baseline operasional untuk tingkat pelanggaran adalah kelompok bangunan tanpa satu pun item sindrom (sindrom-0), yang dijadikan referensi nol; seluruh analisis dosis-respons dinyatakan sebagai rasio terhadap kelompok referensi ini.

4.9 Teknik Pengendalian Ukuran dan Analisis Wilayah

Karena bangunan besar secara mekanis menghasilkan lebih banyak pengaduan dan lebih banyak pelanggaran, ukuran bangunan menjadi variabel pengganggu utama. Pengendaliannya dilakukan dengan dua cara. Pertama, seluruh tingkat pelanggaran dihitung per unit hunian, bukan sebagai jumlah mentah. Kedua, perbandingan dilakukan di dalam pita ukuran yang sama berdasarkan jumlah unit hunian, sehingga bangunan yang dibandingkan setara ukurannya. Ukuran intensitas yang digunakan adalah *pooled rate*, yaitu total pelanggaran dibagi total unit pada satu kelompok, karena ukuran ini tidak runtuh menjadi nol pada kelompok dengan prevalensi rendah.

Untuk RQ2, unit wilayah yang digunakan adalah kode pos (ZIP). Untuk setiap pasangan wilayah dan tahun, dihitung komposisi pengaduan dalam bentuk persentase, bukan jumlah, agar pengelompokan tidak sekadar memisahkan wilayah padat dari wilayah sepi. Selanjutnya digunakan dua lapis fitur. Lapis pertama adalah komposisi mentah. Lapis kedua adalah deviasi, yaitu komposisi wilayah dikurangi komposisi rata-rata seluruh kota pada tahun yang sama. Lapis deviasi inilah yang memisahkan arus perubahan yang menyapu seluruh kota dari karakter khas masing-masing wilayah. Hanya jenis pengaduan yang hidup di seluruh enam tahun yang dipakai sebagai fitur, agar perbandingan antarwaktu tidak terganggu oleh perubahan taksonomi pencatatan. Pendekatan komposisi wilayah ini mengikuti gagasan bahwa struktur jenis pengaduan dapat menjadi tanda tangan khas sebuah lokasi (Wang dkk., 2017).

BAB V

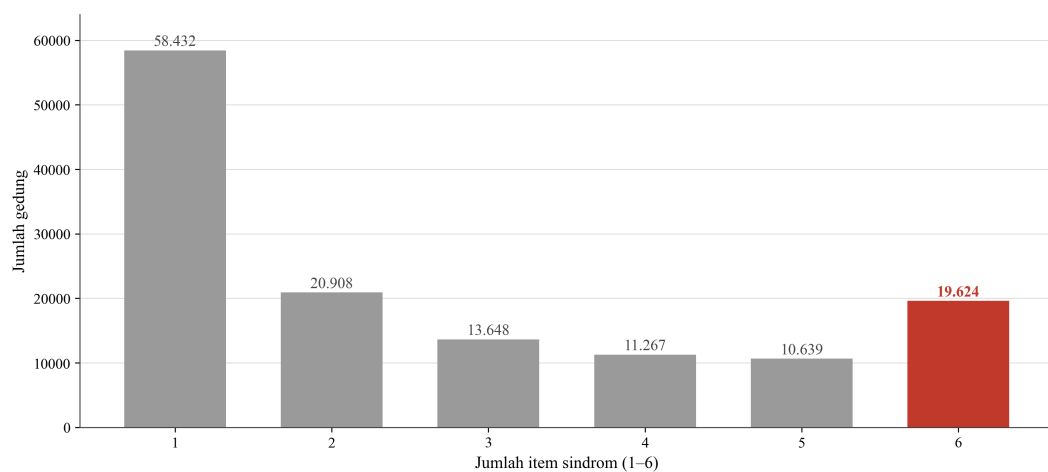
HASIL

5.1 Catatan Pembacaan

Bab ini memuat deskripsi faktual hasil analisis, yaitu angka, distribusi, dan pernyataan observasional, tanpa penafsiran. Seluruh interpretasi, penjelasan sebab, dan kalibrasi klaim disajikan secara terpisah pada Bab VI. Pemisahan ini sengaja dilakukan untuk menjaga batas tegas antara apa yang teramati dan apa yang ditafsirkan.

5.2 Distribusi Sindrom (RQ1)

Distribusi jumlah jenis pengaduan sindrom per bangunan ditampilkan pada Gambar 7. Jumlah bangunan menurun dari satu ke lima item, lalu naik kembali pada enam item. Secara absolut, kelompok dengan enam item memuat 19.624 bangunan, lebih besar daripada kelompok empat item (11.267) maupun lima item (10.639). Kelompok dengan empat item ke atas berjumlah 41.530 bangunan.

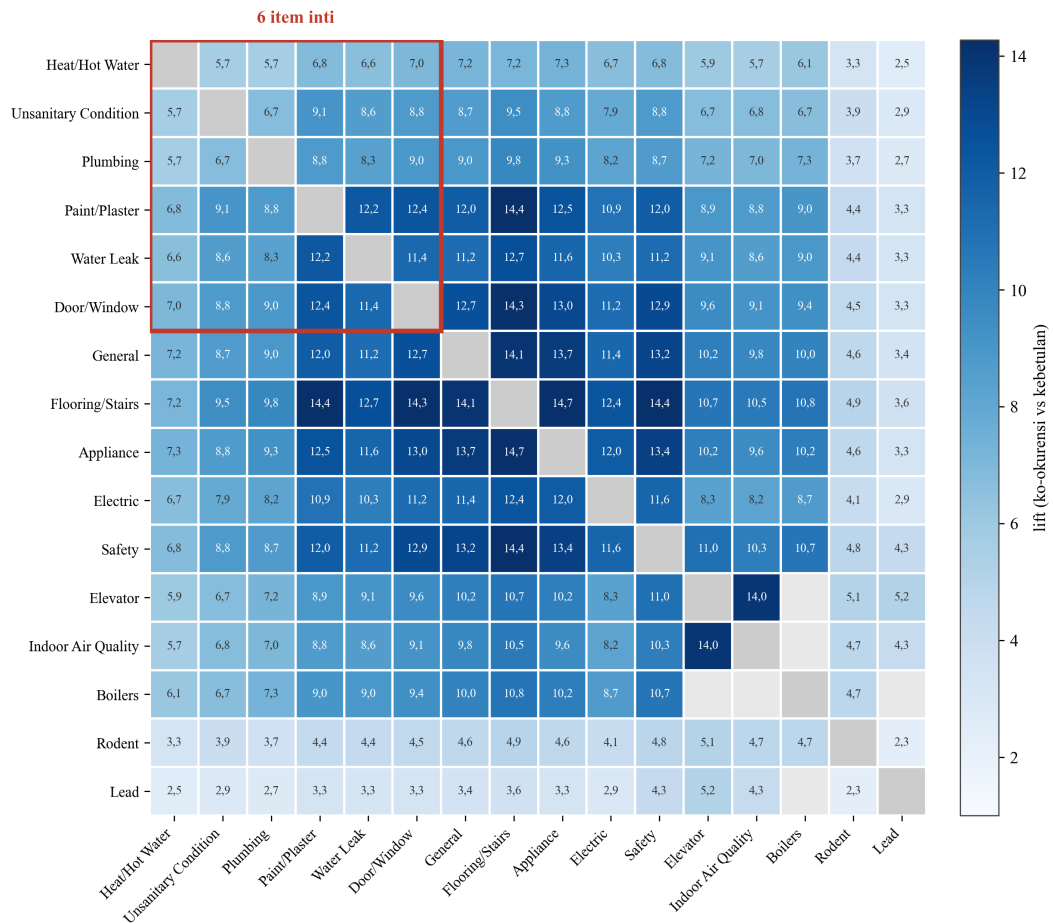


Gambar 7 Jumlah bangunan menurut banyaknya item sindrom interior yang dilaporkan

5.3 Struktur Keterkaitan Antarpengaduan

Penambangan asosiasi menghasilkan 95.902 kombinasi sering. Gambar 8 menampilkan nilai *lift* antarpasangan dari enam belas jenis pengaduan keluarga interior. Untuk enam item inti, rata-rata *lift* dari lima belas pasangannya adalah 8,48 (bootstrap 95% CI [7,44, 9,55] dari 10.000 *resample*, *seed* 42). Pasangan terkuat di antara keenam item inti adalah Cat/Plester dan Pintu/Jendela dengan *lift* 12,40. Perlu dicatat secara jujur bahwa aturan dengan *lift* tertinggi pada keseluruhan data bukanlah sindrom interior, melainkan sekelompok pengaduan layanan makanan

komersial (*Food Establishment, Food Poisoning, dan Noise - Commercial*) dengan *lift* berkisar 22 hingga 33 (lihat Lampiran C). Sindrom interior dibedakan bukan oleh *lift* tertinggi, melainkan oleh frekuensi (*support*) yang tinggi dan, sebagaimana ditunjukkan kemudian, oleh kaitannya dengan pelanggaran terverifikasi.



Gambar 8 Struktur ko-okurensi keluhan kerusakan interior berdasarkan nilai lift

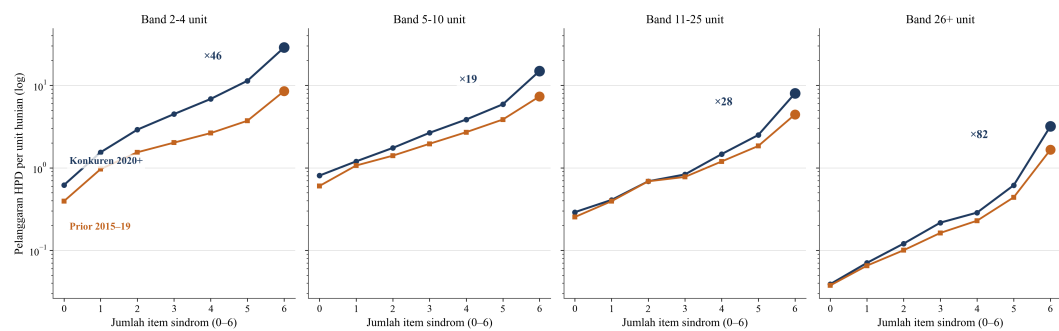
5.4 Hubungan Dosis-Respons dengan Pelanggaran

Gambar 9 menampilkan *pooled rate* pelanggaran HPD per unit menurut jumlah item sindrom, dipisah per pita ukuran bangunan dan per jendela waktu. Pada setiap pita ukuran, nilai naik secara monoton dari sindrom-0 ke sindrom-6. Pada jendela konkuren, rasio antara sindrom-6 dan sindrom-0 berkisar 19 hingga 82 kali bergantung pita ukuran. Pada jendela prior, kenaikan monoton yang sama tetap teramati, dengan rasio sindrom-6 terhadap sindrom-0 berkisar 12 hingga 44 kali bergantung pita ukuran. Tabel 7 merangkum nilai kunci beserta jumlah bangunan tiap kelompok.

Tabel 7 Ringkasan rasio pelanggaran sindrom-6 terhadap sindrom-0 per pita ukuran

Pita ukuran	Rasio konkuren	Rasio prior	Catatan jumlah
Band 2–4 unit	×46	×22	n besar pada sindrom rendah
Band 5–10 unit	×19	×12	sebaran merata
Band 11–25 unit	×28	×18	sebaran merata
Band 26+ unit	×82	×44	n kecil pada sindrom tinggi

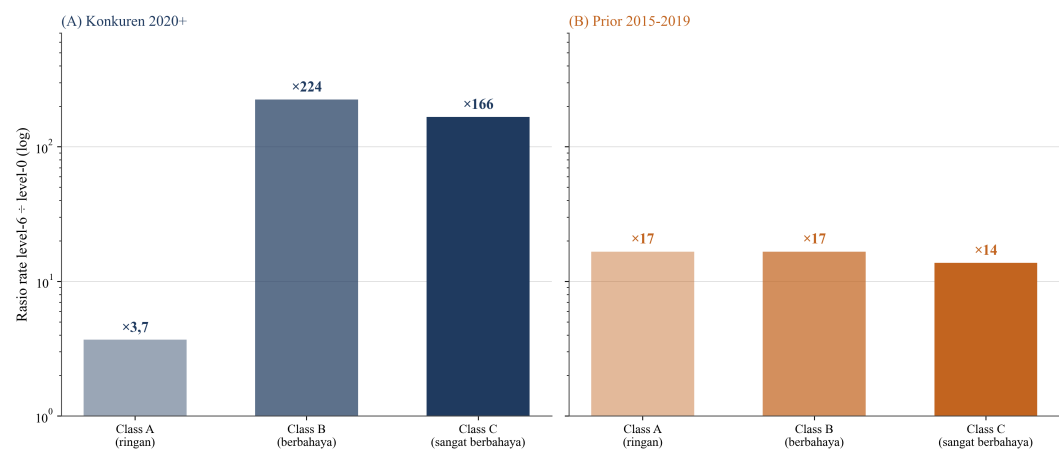
Sumber: building_master_hpd, diolah



Gambar 9 Tingkat pelanggaran HPD per unit menurut jumlah item sindrom di setiap pita ukuran dan kedua jendela waktu

5.5 Hasil Spesifisitas Keparahan

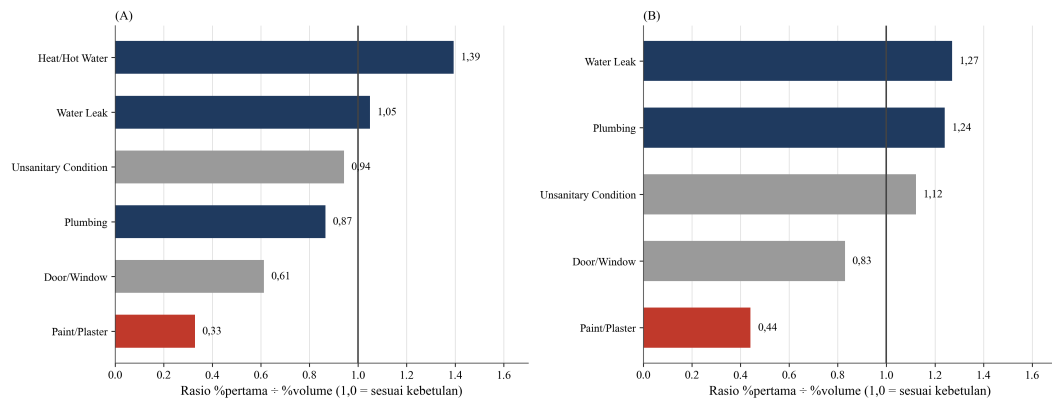
Gambar 10 menampilkan rasio pelanggaran sindrom-6 terhadap sindrom-0 menurut kelas keparahan pada kedua jendela. Pada jendela konkuren, kelas berbahaya dan sangat berbahaya menunjukkan rasio sekitar 224 dan 166 kali, sedangkan kelas ringan sekitar 4 kali. Pada jendela prior, ketiga kelas keparahan menunjukkan rasio yang setara, yaitu sekitar 14 sampai 17 kali.



Gambar 10 Rasio pelanggaran per kelas keparahan pada jendela konkuren dan jendela prior

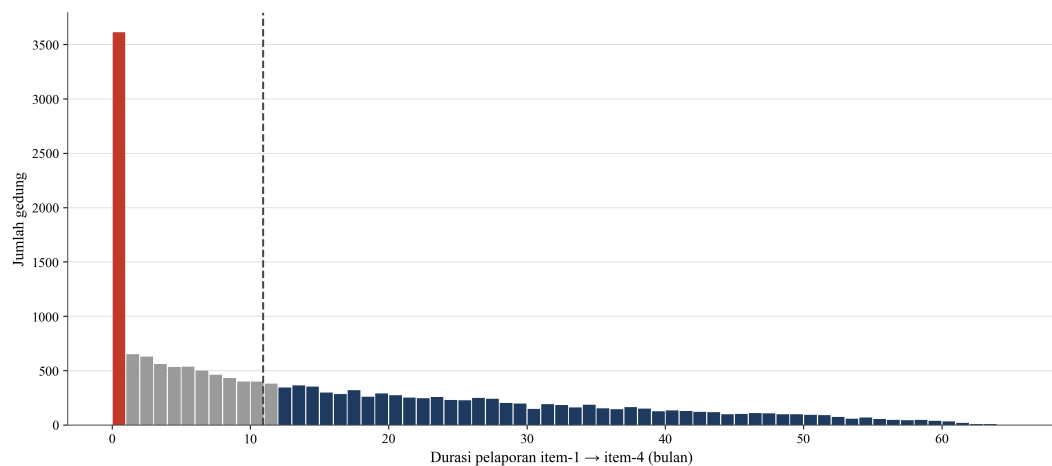
5.6 Urutan dan Kecepatan Temporal

Gambar 11 menampilkan rasio antara persentase suatu item menjadi yang pertama dilaporkan terhadap porsi volumenya. Item terkait air, yaitu Pemanas dan Kebocoran Air, memiliki rasio di atas satu, sedangkan Cat/Plester memiliki rasio sekitar 0,33 hingga 0,44. Pola ini bertahan pada panel sensitivitas tanpa Heat.



Gambar 11 Item yang pertama dilaporkan dinyatakan sebagai rasio terhadap ekspektasi volume

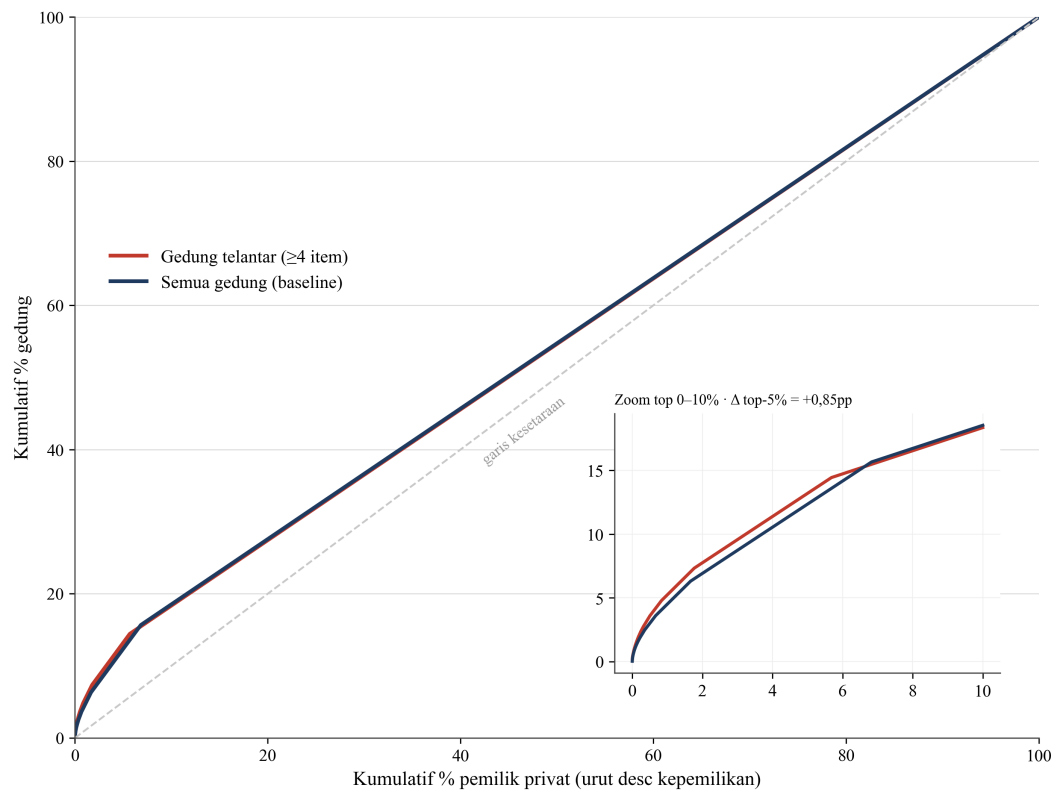
Gambar 12 menampilkan distribusi durasi dari item pertama ke item keempat. Distribusi bersifat dua puncak. Sebanyak 20,8 persen bangunan mencapai durasi kurang dari satu bulan, sementara 47,6 persen melampaui dua belas bulan. Median durasi adalah 332 hari. Sebanyak 43 persen bangunan melaporkan dua atau lebih item pada hari yang sama.



Gambar 12 Distribusi kecepatan akumulasi sindrom yang bersifat dua puncak

5.7 Konsentrasi Kepemilikan

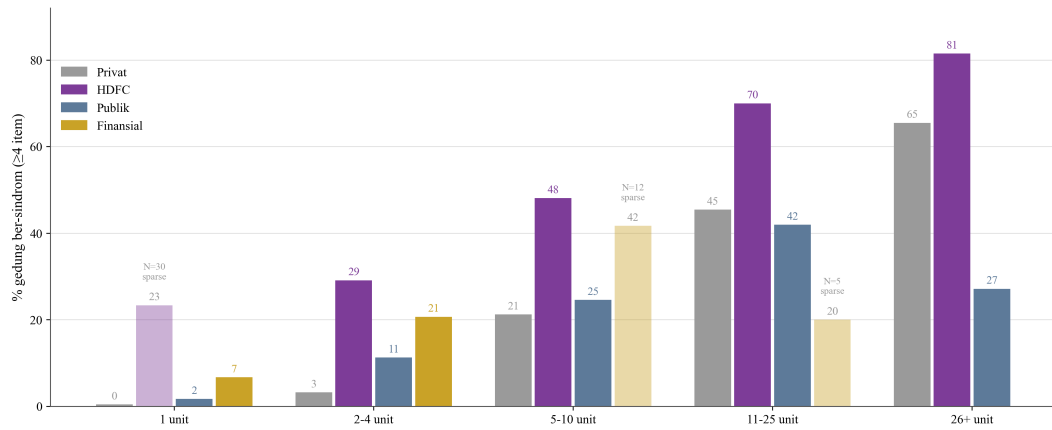
Gambar 13 menampilkan kurva konsentrasi kepemilikan untuk bangunan terabaikan dan baseline. Kedua kurva hampir berimpit, dengan selisih pada lima persen teratas sebesar 0,85 poin persen. Terdapat 85 pemilik yang seluruh portofolionya tergolong ber-sindrom.



Gambar 13 Konsentrasi kepemilikan gedung terabaikan dibandingkan baseline kepemilikan umum

5.8 Sindrom menurut Kategori Pemilik

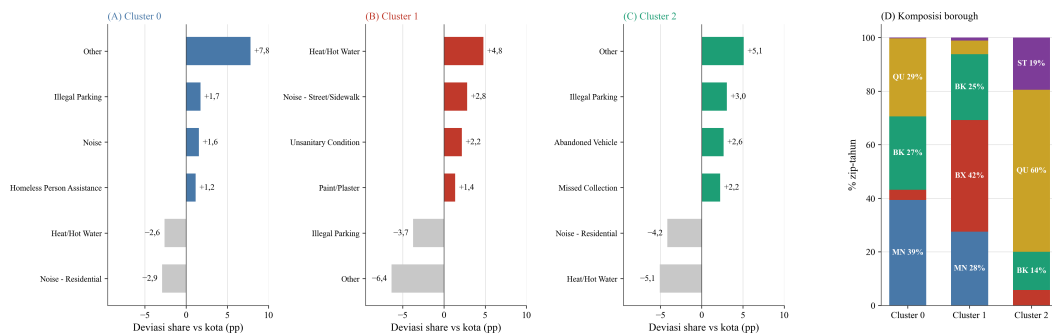
Gambar 14 menampilkan persentase bangunan ber-sindrom per kategori pemilik di tiap pita ukuran. Pada seluruh pita, kategori HDFC memiliki persentase lebih tinggi daripada kategori privat, dengan selisih 16 hingga 27 poin persen. Pada pita 1 unit, jumlah sampel HDFC kecil ($n=30$).



Gambar 14 Persen gedung ber-sindrom per kategori pemilik dengan kontrol ukuran gedung

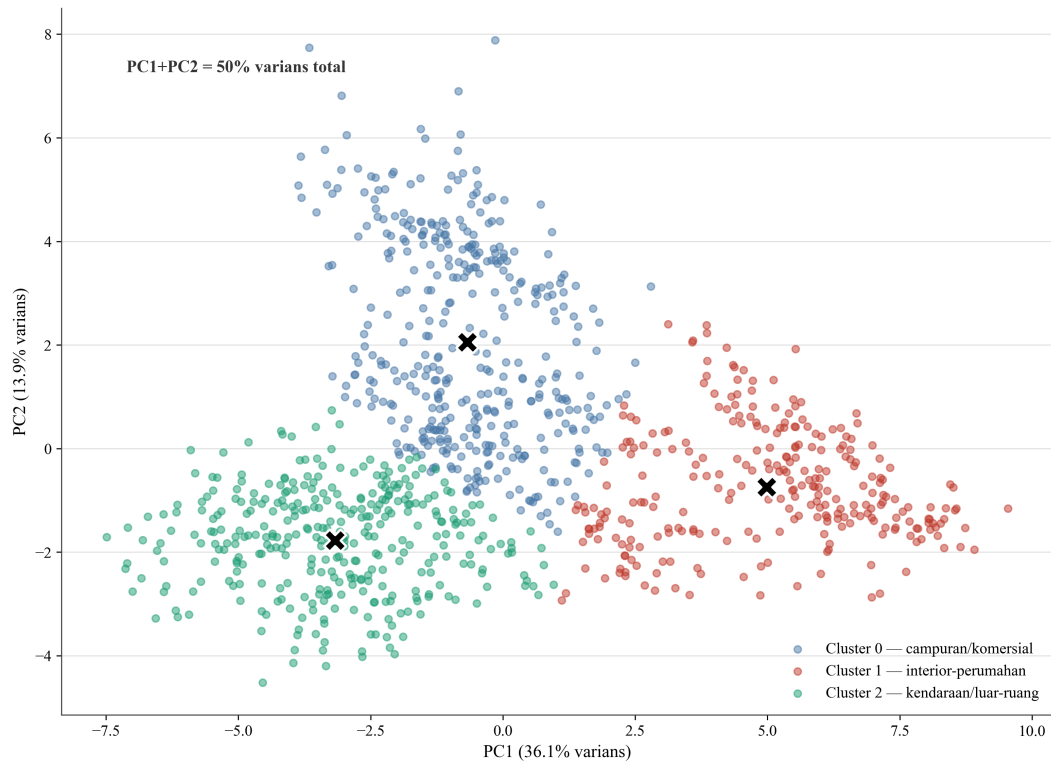
5.9 Tipologi Wilayah (RQ2)

Pengelompokan menghasilkan tiga cluster dengan rata-rata *silhouette* 0,217 (standar deviasi antar-seed 0,0001 dari 20 seed). Gambar 15 menampilkan item paling khas dan komposisi borough tiap cluster. Cluster 0 dicirikan pengaduan campuran-komersial dan terpusat di Manhattan; Cluster 1 dicirikan pengaduan interior-perumahan dan terpusat di Bronx; Cluster 2 dicirikan pengaduan kendaraan dan luar-ruang serta terpusat di Queens dan Staten Island.



Gambar 15 Tipologi wilayah berupa item paling khas per cluster dan komposisi borough

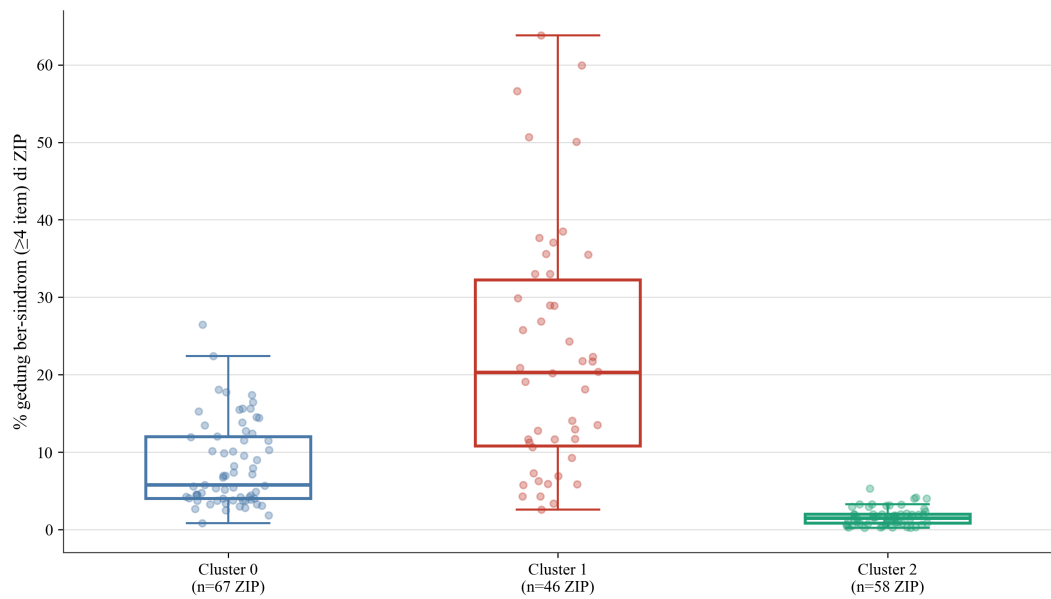
Gambar 16 menampilkan proyeksi PCA tiga cluster. Dua komponen utama menjelaskan 50 persen varians (PC1 36,1 persen; PC2 13,9 persen). Gerombolan ketiga cluster saling bersinggungan pada batasnya.



Gambar 16 Wujud tiga tipe wilayah dalam ruang PCA

5.10 Keterkaitan Lintas-Skala

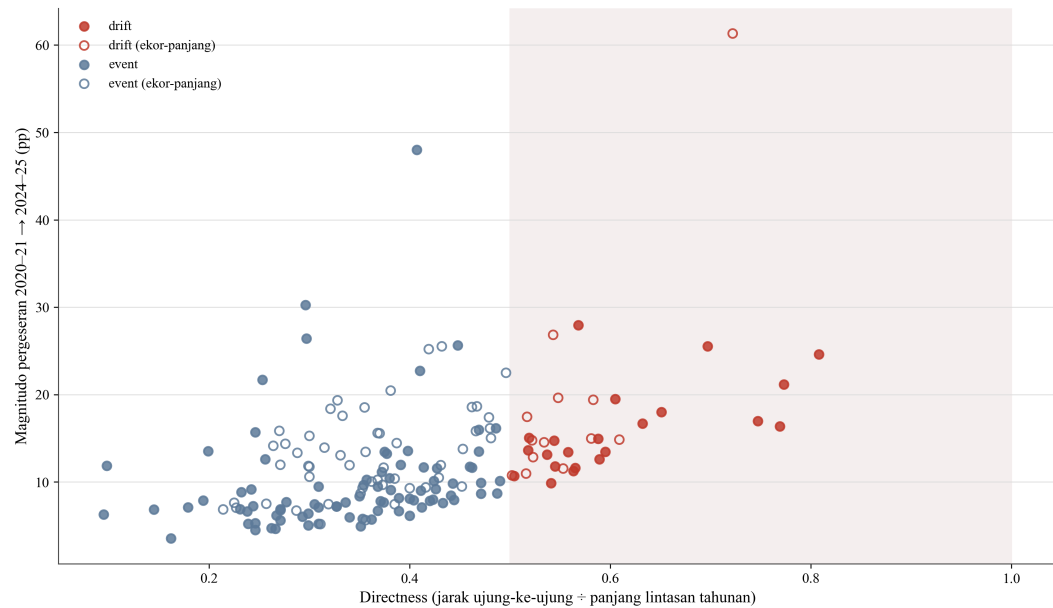
Gambar 17 menampilkan konsentrasi bangunan ber-sindrom per tipe cluster. Median konsentrasi adalah 20,3 persen pada Cluster 1, 6,7 persen pada Cluster 0, dan 1,5 persen pada Cluster 2.



Gambar 17 Konsentrasi gedung ber-sindrom RQ1 menurut tipe cluster wilayah RQ2

5.11 Kestabilan dan Pergeseran Wilayah

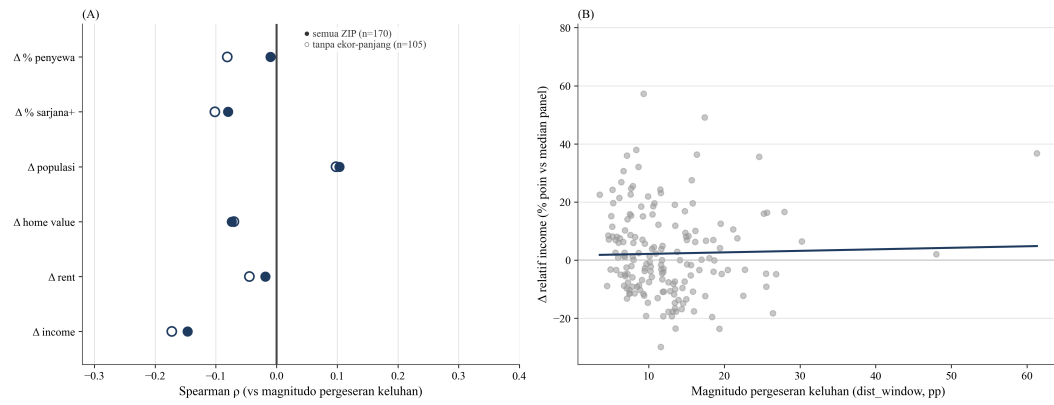
Gambar 18 memisahkan wilayah yang bergeser arah konsisten dari yang berfluktuasi. Sebanyak 84,2 persen wilayah tidak berpindah cluster selama enam tahun. Dari pergeseran yang ada, sekitar sepertiga magnitudonya merupakan arus se-kota. Sejumlah kecil wilayah, terutama koridor di Brooklyn, menunjukkan pergeseran arah yang konsisten.



Gambar 18 Magnitudo dibandingkan konsistensi arah pergeseran komposisi keluhan antarwilayah

5.12 Keselarasan dengan Sosio-Ekonomi

Gambar 19 menampilkan hasil uji keselarasan antara pergeseran pengaduan dan perubahan sosio-ekonomi. Seluruh korelasi Spearman berada pada rentang $|\rho| < 0,18$ dan tidak ada yang signifikan setelah mempertimbangkan dua belas pengujian. Korelasi terkuat adalah untuk pendapatan, yaitu $\rho \approx -0,15$ dengan $p \approx 0,057$.



Gambar 19 Hasil uji korelasi pergeseran karakter keluhan dengan perubahan sosio-ekonomi

BAB VI

PEMBAHASAN KRITIS

6.1 Interpretasi Temuan Utama

Bagian ini menafsirkan hasil Bab V dengan klaim yang dikalibrasi secara epistemik. Untuk sindrom enam item, hasil yang teramati konsisten dengan interpretasi bahwa sindrom merupakan penanda pengabaian kronis. Penafsiran ini diperkuat oleh fakta bahwa asosiasi dengan pelanggaran HPD telah lolos dua pengujian pengendalian, yaitu ukuran bangunan melalui stratifikasi dan sirkularitas melalui jendela prior. Kenaikan jumlah bangunan pada enam item, dibandingkan empat atau lima, konsisten dengan gagasan bahwa ketika sebuah bangunan terabaikan, banyak sistem cenderung rusak sebagai satu kesatuan, bukan satu per satu. Klaim ini berlaku pada dataset ini dan tidak diajukan sebagai hukum umum.

Untuk pola temporal, distribusi dua puncak konsisten dengan interpretasi bahwa pengabaian menjadi terlihat melalui dua jalur, yaitu kerusakan yang menumpuk dalam diam lalu dilaporkan sekaligus saat ada pemicu, dan akumulasi bertahap selama bertahun-tahun. Urutan pelaporan yang menempatkan item air lebih awal dan cat/plester paling akhir konsisten dengan interpretasi cat/plester sebagai indikator yang tertinggal. Perlu ditegaskan bahwa ini adalah urutan pelaporan, bukan urutan sebab-akibat fisik.

Untuk tipologi wilayah, hasil yang teramati konsisten dengan interpretasi karakter struktural yang lengket, bukan cerminan kejadian sesaat, mengingat 84,2 persen wilayah tidak berpindah cluster selama enam tahun. Untuk temuan HDFC, selisih yang bertahan setelah pengendalian ukuran konsisten dengan interpretasi bahwa stok perumahan terjangkau merupakan titik kumpul beban pengabaian. Penafsiran ini sengaja tidak menyatakan arah kausalitas.

6.2 Uji Ketangguhan Interpretasi

Untuk setiap klaim utama dibangun argumen tandingan terkuat, lalu dinilai apakah data mendukungnya.

Klaim pertama yang dipertahankan adalah bahwa sindrom merupakan penanda pengabaian nyata, bukan artefak perilaku melapor. Interpretasi alternatif yang sama-sama konsisten secara logis adalah bahwa warga di wilayah tertentu cenderung melaporkan lebih banyak jenis masalah tanpa kondisi bangunan yang benar-benar lebih buruk. Interpretasi utama dipertahankan karena jendela prior memperlihatkan pelanggaran HPD sudah tinggi sebelum periode pengaduan, dan asosiasi bertahan di

seluruh borough. Namun diakui bahwa argumen tandingan tidak dapat sepenuhnya disingkirkan untuk komunitas dengan pola pelaporan yang sangat rendah.

Klaim kedua adalah bahwa beban pengabaian menumpuk pada HDFC secara struktural. Argumen tandingan adalah bahwa bangunan HDFC mungkin lebih tua atau lebih besar sehingga menghasilkan lebih banyak pengaduan secara mekanis. Interpretasi utama dipertahankan karena analisis dilakukan di dalam pita ukuran yang sama dan dinormalisasi per unit. Namun diakui bahwa data tidak dapat memisahkan apakah kondisi buruk mendahului atau justru mengikuti perubahan status menjadi HDFC, sehingga arah kausalitas tetap terbuka.

Klaim ketiga adalah bahwa tipologi wilayah merupakan sinyal tersendiri yang tidak mencerminkan gentrifikasi. Argumen tandingan adalah bahwa korelasi nol dengan ACS dapat muncul karena indikator ACS yang dipilih kurang tepat, atau karena ada jeda waktu antara perubahan sosio-ekonomi dan perubahan pengaduan. Argumen tandingan ini dinilai valid dan diperlakukan sebagai limitasi yang diakui, bukan disingkirkan.

6.3 Pemeriksaan Variabel Pengganggu

Tabel 8 merangkum variabel pengganggu utama, cara pengendaliannya, dan sisa kekhawatiran yang tetap ada.

Tabel 8 Pemeriksaan variabel pengganggu dan sisa kekhawatiran

Variabel pengganggu	Cara dikontrol	Sisa kekhawatiran
Ukuran bangunan	Stratifikasi pita ukuran dan normalisasi per unit	Pita ukuran masih heterogen di dalamnya
Lokasi geografis	Dekomposisi deviasi terhadap rata-rata kota	Efek hiperlokal tidak tertangkap
Perilaku inspektur HPD	Tidak dikontrol	Inspektur mungkin lebih sering mengunjungi area tertentu secara tidak acak; limitasi yang tidak dapat dieliminasi
Bias pelaporan warga	Sebagian melalui jendela prior	Komunitas dengan pelaporan rendah mungkin tetap kurang terwakili

Sumber: analisis peneliti

6.4 Ukuran Efek versus Signifikansi Statistik

Dengan 764.114 keranjang bangunan, hampir semua perbedaan akan lolos uji signifikansi statistik bahkan ketika perbedaannya kecil secara praktis. Oleh karena itu, penilaian difokuskan pada ukuran efek, bukan sekadar signifikansi. Rasio pelanggaran HPD pada jendela prior sebesar 12 hingga 44 kali antara sindrom-6 dan sindrom-0, bergantung pada ukuran, merupakan ukuran efek yang besar dan bermakna praktis untuk penentuan prioritas inspeksi, karena menunjukkan bahwa bangunan ber-sindrom lengkap memikul beban pelanggaran terverifikasi yang berkali-kali lipat. Secara absolut, kelompok sindrom-4 ke atas mencakup 41.530 bangunan, sehingga intervensi yang menyasar kelompok ini menjangkau populasi bangunan yang substansial.

Untuk temuan HDFC, selisih 16 hingga 27 poin persen merupakan ukuran efek yang besar dalam konteks prevalensi sindrom, sehingga cukup untuk membenarkan perhatian kebijakan khusus pada stok perumahan terjangkau, dengan catatan arah kausalitas yang belum tuntas. Sebaliknya, untuk hasil ACS, ukuran efek yang kecil ($|\rho| < 0,18$) memperkuat kesimpulan bahwa tidak ada keselarasan yang bermakna praktis, sehingga signifikansi marginal pada satu indikator pun tidak mengubah kesimpulan.

6.5 Batas Generalisasi

Tabel 9 menyatakan secara eksplisit di mana temuan berlaku dan di mana perlu verifikasi sebelum diterapkan.

Tabel 9 Batas generalisasi temuan

Berlaku untuk	Perlu verifikasi sebelum diterapkan ke
Bangunan perumahan NYC yang terdaftar dalam sistem 311 dan HPD pada periode 2020–2024	Kota lain dengan sistem pengaduan dan inspeksi yang berbeda
Bangunan dengan BBL teridentifikasi yang masuk unit analisis	Periode sebelum 2020, karena anomali pandemi dapat mengubah pola pelaporan
Pola asosiasi pada jenis pengaduan yang stabil sepanjang enam tahun	Bangunan tanpa BBL terdaftar atau stok bangunan informal
Validasi yang bersandar pada catatan pelanggaran terverifikasi	Kota tanpa sistem pelanggaran terverifikasi sebagai validator independen

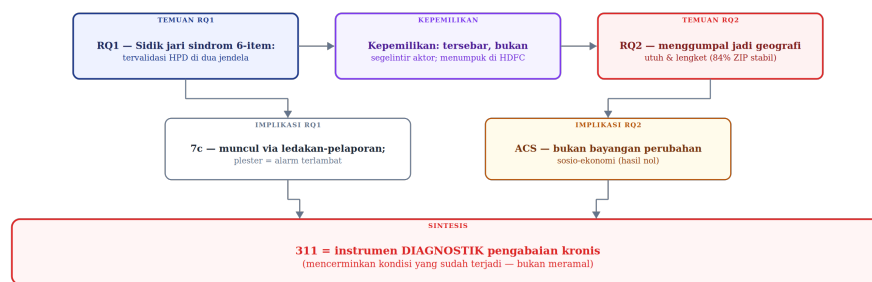
Sumber: analisis peneliti

BAB VII

PENUTUP

7.1 Kesimpulan

Sintesis seluruh temuan ditampilkan pada Gambar 20. Penelitian ini menjawab kedua pertanyaan penelitian secara langsung.



Gambar 20 Sintesis temuan mengenai apa yang dikatakan data 311 tentang pengabaian bangunan

RQ1 menanyakan apakah jenis-jenis pengaduan tertentu muncul bersama melebihi kebetulan sehingga membentuk sidik jari pengabaian bangunan. Jawaban penelitian ini adalah ya, dengan bukti berikut: terbentuk sindrom enam item dengan rata-rata *lift* antarpasangan inti 8,48 kali di atas kebetulan (terkuat 12,40); bangunan cenderung memuat keenam item secara lengkap; *pooled rate* pelanggaran HPD naik monoton seiring kelengkapan sindrom di setiap pita ukuran; dan kenaikan ini bertahan pada jendela prior yang independen. Hubungan ini berlaku pada bangunan perumahan NYC periode 2020–2024, dengan catatan bahwa data mencerminkan pelaporan dan bukan seluruh kondisi fisik. Analisis temporal juga menunjukkan item air cenderung dilaporkan lebih awal dan cat/plester tertinggal.

RQ2 menanyakan apakah wilayah berbeda dalam komposisi pengaduan dan apakah komposisi itu bergeser sepanjang waktu. Jawaban penelitian ini adalah ya untuk perbedaan dan sebagian besar tidak untuk pergeseran, dengan bukti berikut: terbentuk tiga tipologi wilayah yang bergradasi dengan *silhouette* 0,22; karakter wilayah lengket dengan 84,2 persen wilayah tidak berpindah; dan pergeseran karakter pengaduan tidak menunjukkan korelasi berarti dengan perubahan sosio-ekonomi. Klaim bahwa karakter pengaduan sepenuhnya independen dari sosio-ekonomi tidak dapat ditegakkan secara mutlak, melainkan dinyatakan sebagai tidak ditemukannya keselarasan pada dataset dan indikator yang diuji.

Secara keseluruhan, data 311 memuat sinyal yang terverifikasi dan konsisten de-

ngan kondisi pengabaian bangunan yang bersifat kronis. Dengan catatan bahwa data mencerminkan kondisi yang telah terjadi dan bukan meramalkan masa depan, data 311 berpotensi menjadi instrumen diagnostik untuk penentuan prioritas perhatian terhadap perumahan.

7.2 Keterbatasan

Beberapa keterbatasan perlu disampaikan secara jujur. Pertama, data 311 merekam pelaporan dan bukan kondisi sebenarnya, sehingga perubahan dapat mencerminkan kondisi maupun perilaku melapor (Kontokosta dan Hong, 2021). Kedua, unit analisis adalah lot pajak, bukan bangunan fisik. Ketiga, analisis konsentrasi kepemilikan merupakan batas bawah karena entitas berbeda nama dapat memiliki pemilik yang sama. Keempat, kode pos merupakan artefak administratif, bukan batas wilayah sosial. Kelima, perilaku inspektur HPD tidak dapat dikontrol dan tetap menjadi sumber ketidakpastian. Keenam, seluruh analisis bersifat deskriptif dan tidak mengandung klaim sebab-akibat.

7.3 Saran

Saran disusun dalam tiga kategori sesuai tingkat kebutuhan sumber dayanya.

Pertama, saran yang dapat ditindaklanjuti langsung dengan data yang sudah tersedia. Penelitian lanjutan dapat menganalisis kesesuaian isi antara jenis pengaduan dan jenis pelanggaran HPD menggunakan dataset *building_master_hpd* yang sudah terbentuk, serta memperdalam analisis koridor Brooklyn yang menunjukkan pergeseran arah konsisten.

Kedua, saran yang membutuhkan data tambahan. Pemisahan antara penjelasan pendanaan dan penjelasan seleksi pada temuan HDFC memerlukan data tanggal konversi kepemilikan, yang dapat diperoleh dari NYC Department of Finance. Pene-lusuran kepemilikan sesungguhnya untuk mengatasi fragmentasi nama memerlukan data *beneficial ownership* tambahan.

Ketiga, saran yang membutuhkan desain studi berbeda. Pertanyaan mengenai arah kausalitas, baik pada sindrom maupun pada HDFC, tidak dapat dijawab dengan data observasional dan memerlukan desain kuasi-eksperimental, misalnya memanfaatkan perubahan kebijakan sebagai sumber variasi. Deteksi anomali pelaporan untuk membedakan kondisi nyata dari perilaku melapor memerlukan model dasar dan karenanya mendekati ranah *machine learning* yang berada di luar lingkup penelitian ini.

LAMPIRAN

Lampiran A — Spesifikasi Lingkungan Komputasi

Komponen	Spesifikasi
Bahasa	Python 3.12.8
Pengolahan data	polars 1.41.2 (mesin utama); pandas 3.0.3 (hanya untuk mlxtend)
Komputasi numerik	numpy 2.4.6; scipy 1.17.1
Penambangan asosiasi	mlxtend 0.24.0 (implementasi FP-Growth)
Pengelompokan	scikit-learn 1.9.0 (KMeans, silhouette_score)
Visualisasi	matplotlib 3.10.9
Sistem operasi	Windows 11 (build 10.0.26200)
Prosesor	AMD Ryzen (Zen 3, AMD64 Family 25), 12 core logis
Memori	33,5 GB
<i>Random seed</i>	42 (inisialisasi K-means dan <i>bootstrap</i>)
Repositori	https://github.com/Ikrar06/nyc-311-data-mining

Lampiran B — Log Keputusan Preprocessing

Langkah	Keputusan dan alasan	Sebelum	Sesudah
Muat data mentah	Baca seluruh catatan 311	21.340.927	21.340.927
Normalisasi ComplaintType	Samakan huruf besar/kecil untuk hapus duplikat semu antar-label	21.340.927	21.340.927
Periksa duplikat UniqueKey	0 duplikat ditemukan; tidak ada baris dibuang	21.340.927	21.340.927
Perbaiki LotArea PLUTO	Tangani pemisah ribuan agar luas benar	—	—
Saring BBL tidak valid	Buang BBL kosong (2.486.694) dan BBL=0 (2.393); BBL kunci unit analisis	21.340.927	18.851.840
Bentuk keranjang	Agregasi jenis pengaduan unik per BBL	18.851.840	764.114

Lampiran C — Tabel Hasil Lengkap

Tabel C.1 menyajikan lima aturan asosiasi dengan *lift* tertinggi pada keseluruhan data. Sebagaimana dibahas pada Bab V, aturan ber-*lift* tertinggi berasal dari klaster layanan makanan komersial, bukan dari sindrom pemeliharaan interior.

Antecedent	Consequent	Supp.	Conf.	Lift
Consumer Complaint, Food Establishment	Food Poisoning	0,0058	0,395	33,07
Food Poisoning	Consumer Complaint, Food Establishment	0,0058	0,488	33,07
Food Establishment, Noise - Commercial	Food Poisoning	0,0050	0,367	30,68
Food Poisoning	Food Establishment, Noise - Commercial	0,0050	0,418	30,68
Food Establishment, Noise - Street/Sidewalk	Food Poisoning	0,0055	0,349	29,16

Catatan: lima belas aturan *lift*-tertinggi berikutnya (peringkat 6–20) merupakan varian dari klaster yang sama (*Food Establishment*, *Food Poisoning*, *Outdoor Dining*) dengan *lift* berkisar 22,5–28,7. Sebagai pembandingan, lima belas pasangan dari enam item sindrom interior memiliki rata-rata *lift* 8,48 (95% CI [7,44, 9,55]) dengan pasangan terkuat, Cat/Plester–Pintu/Jendela, sebesar 12,40. Sindrom interior karenanya dibedakan oleh frekuensi dan validasi pelanggaran, bukan oleh *lift* tertinggi.

Tabel C.2 menyajikan perbandingan performa pengelompokan untuk beberapa nilai k .

k	Silhouette	Inertia	Interpretabilitas
2	0,2661	27.241,7	<i>Silhouette</i> tertinggi, tetapi menggabung tipe berbeda
3	0,2167	23.326,8	Dipilih: tiga tipe interpretatif
4	0,2100	21.084,1	Muncul cluster kecil tak interpretatif
5	0,1733	19.855,5	Fragmentasi tanpa makna domain

Lampiran D — Log Eksperimen yang Tidak Dipilih

Tiga konfigurasi berikut dicoba namun tidak dipilih untuk hasil utama.

1. Konfigurasi *min_support* 0,01 yang lebih tinggi diuji, namun memangkas jumlah kombinasi secara tajam dari 95.902 menjadi 16.454, sehingga sebagian pola yang masih relevan secara domain (termasuk beberapa pasangan inti pada subpopulasi kecil) ikut hilang, sehingga ditolak.
2. Pengelompokan dengan $k = 4$ diuji. Nilai *silhouette*-nya (0,2100) hanya sedikit di bawah $k = 3$ (0,2167), tetapi $k = 4$ menghasilkan satu cluster berukuran sangat kecil yang tidak dapat diinterpretasikan secara domain, sehingga ditolak demi tipologi yang lebih bermakna.
3. Pengelompokan menggunakan data komposisi mentah tanpa dekomposisi deviasi diuji, namun hanya memisahkan wilayah padat dari wilayah sepi dan gagal menangkap karakter khas pengaduan, sehingga ditolak demi lapis deviasi.

DAFTAR PUSTAKA

- Hashad, A. A., Khaw, K. W., Alnoor, A., & Chew, X. (2024). Exploratory analysis with association rule mining algorithms in the retail industry. *Malaysian Journal of Computing*, 9(1), 1746–1758. <https://doi.org/10.24191/mjoc.v9i1.21433>
- Kontokosta, C. E., & Hong, B. (2021). Bias in smart city governance: How socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions. *Sustainable Cities and Society*, 64, 102503. <https://par.nsf.gov/servlets/purl/10314387>
- Li, S., & Hyder, A. (2020). 311 service requests as indicators of neighborhood distress and opioid use disorder. *Scientific Reports*, 10, 19579. <https://doi.org/10.1038/s41598-020-76685-z>
- New York City Department of City Planning. (2024). *Primary Land Use Tax Lot Output (PLUTO) data dictionary*. NYC Open Data. <https://www.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page>
- New York City Department of Housing Preservation and Development. (2024). *Housing maintenance code violations*. NYC Open Data. <https://data.cityofnewyork.us/Housing-Development/Housing-Maintenance-Code-Violations/wvxf-dwi5>
- New York City Open Data. (2024). *311 service requests from 2010 to present*. <https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9>
- Pazdora, G. M., Wang, T., Leung, C. K., & Chauhan, A. S. (2022). Predictive big data analytics for service requests: A framework. *Procedia Computer Science*, 198, 102–111. <https://doi.org/10.1016/j.procs.2021.12.216>
- U.S. Census Bureau. (2024). *American Community Survey 5-year data (2015–2019 dan 2020–2024)*. <https://www.census.gov/data/developers/data-sets/acs-5year.html>
- Wang, L., Qian, C., Kats, P., Kontokosta, C., & Sobolevsky, S. (2017). Structure of 311 service requests as a signature of urban location. *PLOS ONE*, 12(10), e0186314. <https://doi.org/10.1371/journal.pone.0186314>