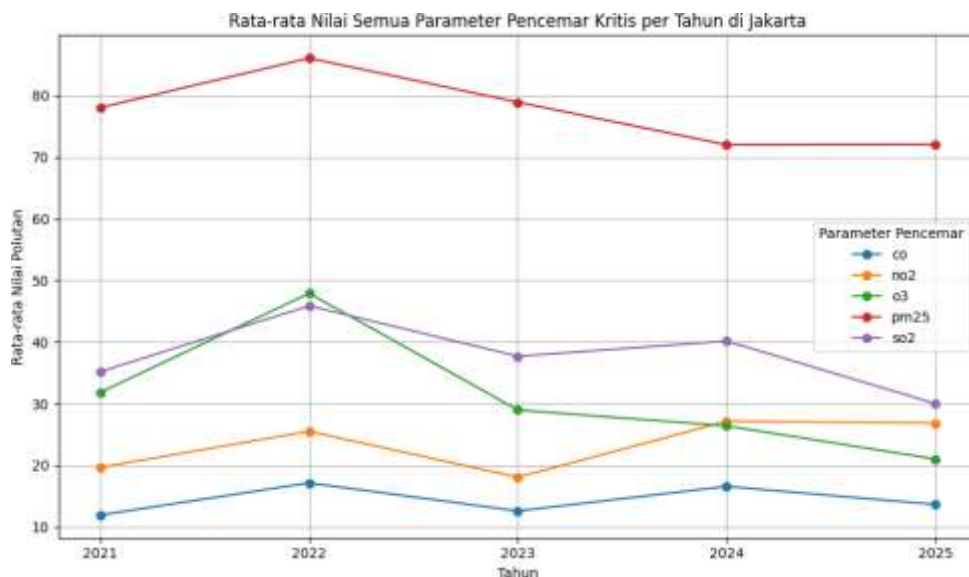


BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurunnya kualitas udara disebabkan oleh polusi udara [1]. Polusi udara adalah kontaminasi udara oleh zat-zat berbahaya seperti partikel halus ($PM_{2.5}$ dan PM_{10}), karbon monoksida (CO), sulfur dioksida (SO_2), nitrogen dioksida (NO_2), ozon (O_3), dan senyawa organik volatil (VOC) yang berasal dari kendaraan bermotor, aktivitas industri, pembakaran terbuka, dan emisi rumah tangga [2]. Dinamika polutan tidak hanya dipengaruhi oleh sumber emisi, tetapi juga sangat dipengaruhi oleh kondisi meteorologi seperti suhu, kelembaban relatif, kecepatan angin, curah hujan, dan tekanan atmosfer yang berperan dalam transportasi, dispersi, transformasi kimia, dan deposisi polutan [3]. Masing-masing parameter meteorologi memiliki pengaruh spesifik, misalnya suhu mempengaruhi turbulensi atmosfer dan reaksi kimia [4], kelembaban dapat menurunkan konsentrasi polutan melalui deposisi basah namun juga membentuk aerosol sekunder, dan curah hujan berintensitas tinggi efektif menurunkan konsentrasi $PM_{2.5}$ secara signifikan [5].



Gambar 1. 1 Grafik Kualitas Udara DKI Jakarta

Berdasarkan laporan World Air Quality 2024 dari IQAir, Indonesia

merupakan negara dengan kualitas udara terburuk di Asia Tenggara dan peringkat ke-15 dunia [6]. Sebagai kota metropolitan dengan 10,68 juta jiwa penduduk pada tahun 2025 [7], pesatnya laju pertumbuhan penduduk dan urbanisasi di Jakarta menyebabkan degradasi lingkungan [8], terutama akibat lalu lintas kendaraan yang menyumbang 75% dari total emisi. Data lima tahun terakhir dari Stasiun Pemantau Kualitas Udara (SPKU) Satu Data Jakarta pada Gambar 1 menunjukkan PM2.5 memiliki rata-rata nilai polutan tertinggi dibandingkan parameter pencemar lainnya (CO, NO₂, O₃, dan SO₂), dengan nilai berkisar antara 72–86 per tahun. Konsentrasi PM2.5 mencapai puncaknya sekitar 86 pada tahun 2022, kemudian mengalami penurunan bertahap hingga mencapai sekitar 72 pada tahun 2024 dan 2025. Nilai ini secara konsisten jauh melebihi baku mutu tahunan nasional yang ditetapkan oleh Peraturan Pemerintah (PP) Nomor 22 Tahun 2021 tentang Penyelenggaraan Perlindungan dan Pengelolaan Lingkungan Hidup yaitu 15 µg/m³ per tahun. Partikel PM2.5 adalah polutan yang berukuran sangat kecil sekitar 2,5 mikron, merupakan partikel halus yang dapat terhirup [9].

Pencemaran udara di Jakarta memberikan dampak serius terhadap kesehatan, lingkungan, dan perekonomian. Polusi udara berkontribusi terhadap sekitar 36 kematian per 100.000 jiwa per tahun dan menurunkan rata-rata harapan hidup hingga 1,6 tahun [10]. Paparan pencemaran udara meningkatkan risiko penyakit pernapasan seperti ISPA, asma, bronkitis kronis, dan penyakit kardiovaskular, dengan jumlah kasus ISPA mencapai lebih dari 1,9 juta pada periode Januari–Oktober 2025 [11], [12]. Selain dampak kesehatan, pencemaran udara menimbulkan kerugian ekonomi nasional hingga 2–3% dari PDB per tahun serta menyebabkan degradasi ekosistem dan kerusakan infrastruktur [2], [13].

Berdasarkan hal tersebut untuk meningkatkan upaya dalam pengendalian dan pencegahan polusi udara maka diperlukan penelitian yang dapat mengidentifikasi faktor-faktor yang memengaruhi dinamika kualitas udara. Interaksi kompleks antar polutan dan kondisi meteorologi membentuk pola keterkaitan tertentu yang perlu diidentifikasi menggunakan metode yang mampu mengungkap pola tersebut. Penelitian ini menggunakan metode *association rule mining*. *Association rule* merupakan salah satu teknik data mining yang digunakan untuk menemukan asosiasi, korelasi, pola frekuensi di antara set item dalam basis

data [13]. Cara kerja metode ini yaitu dengan mengidentifikasi itemset yang sering muncul bersama dalam dataset, kemudian membentuk aturan asosiasi dalam bentuk *'if-then'*. Pentingnya sebuah aturan asosiasi dapat ditentukan dari dua tolok ukur, yaitu *support* dan *confidence*. *Support* merupakan nilai yang menggambarkan seberapa sering item atau kombinasi item muncul dalam keseluruhan data. Sedangkan *confidence* adalah ukuran yang menunjukkan seberapa kuat hubungan antara item dalam suatu aturan [14]. Untuk menemukan aturan asosiasi ini menggunakan algoritma *FP-Growth*.

Algoritma *FP-Growth* merupakan tingkatan dari algoritma apriori yang menggunakan alternatif frekuensi itemsets berdasarkan angka yang sering muncul pada setiap transaksi (*frequent itemset*) dalam sebuah kumpulan data. Konsep algoritma *FP-Growth* yaitu pembentukan pohon (*tree*) atau *FP-Tree* dalam pencarian *frequent itemsets*. Dengan menggunakan struktur *FP-Tree*, algoritma ini mampu mengekstraksi frequent itemset tanpa memerlukan proses pembentukan kandidat (*candidate generation*) secara berulang seperti pada algoritma Apriori, sehingga hanya membutuhkan dua kali proses pemindaian (*scan*) terhadap basis data dan menghasilkan waktu komputasi yang lebih cepat dan hemat sumber daya [15].

Penelitian mengenai penerapan Association Rule dan *FP-Growth* telah dilakukan oleh beberapa peneliti sebelumnya pada berbagai bidang kajian. Studi yang dilakukan oleh Wu, Wen, dan Zhu (2024) dengan judul "Association Rule Mining with a Special Rule Coding and Dynamic Genetic Algorithm for Air Quality Impact Factors in Beijing, China" yang dipublikasikan pada PLoS ONE 19(3): e0299865 bertujuan untuk menerapkan association rule mining dengan algoritma genetik dinamis untuk menganalisis faktor-faktor yang mempengaruhi kualitas udara di Beijing. Penelitian ini muncul karena teknik ARM tradisional cenderung fokus pada high frequency terms dan mengabaikan pentingnya low-frequency terms, serta menyebabkan pemborosan ruang penyimpanan yang tidak perlu. Metode yang digunakan adalah Dynamic Genetic Algorithm (DGA) yang dikombinasikan dengan ARM, dengan menerapkan strategi multi-information coding dan dynamic crossover/mutation rate pada dataset kualitas udara Beijing tahun 2016. Temuan penelitian menunjukkan bahwa strategi unit point multi-

information coding berhasil mengurangi konsumsi penyimpanan ruang aturan sebesar 50%, indeks evaluasi baru dapat menambang rules dengan tingkat interest hingga 90% termasuk dari low-frequency items, serta mampu mengurangi waktu pencarian sekitar 20% dibandingkan dengan beberapa algoritma meta-heuristic lain sambil meningkatkan konvergensi[16].

Penelitian oleh Salsabila, Sulistiyowati, dan Padilah (2022) berjudul "Pencarian Pola Pemakaian Obat Menggunakan Algoritma *FP-Growth*" yang dipublikasikan pada Journal of Applied Informatics and Computing (JAIC) Vol. 6, No. 2 bertujuan untuk mencari pola pemakaian obat guna membantu pengelolaan stok obat di sebuah klinik yang berlokasi di Kabupaten Brebes. Latar belakang penelitian ini adalah terjadinya kelebihan stok obat yang mengakibatkan ketidaksesuaian antara jumlah data stok dengan stok yang tersedia secara fisik akibat perencanaan yang kurang tepat. Metodologi yang diterapkan menggunakan framework Knowledge Discovery in Databases (KDD) dengan tahapan data selection, pre-processing, transformation, data mining menggunakan algoritma *FP-Growth*, dan evaluation melalui aplikasi Rapid Miner. Dataset yang digunakan sebanyak 2.071 transaksi penjualan obat periode Januari-Desember 2020, yang setelah melalui proses data cleaning tersisa 1.963 transaksi dan kemudian dikelompokkan menjadi 307 data berdasarkan tanggal transaksi. Hasil penelitian menghasilkan 6 aturan asosiasi dengan nilai minimum support sebesar 75 frekuensi atau 23% dan minimum confidence sebesar 75%. Kombinasi itemset terbesar mencapai 3 item dengan rata-rata nilai lift ratio sebesar 1.267. Item obat yang paling sering muncul dalam aturan asosiasi adalah Alpara Tab dan Fg. Troches yang masing-masing muncul sebanyak 4 kali dalam kombinasi yang berbeda [16], [17].

Studi yang dilakukan Ali, Krisbiantoro, dan Afiana (2025) dengan judul "Penerapan Rekomendasi Algoritma *FP-Growth* Untuk Tren Penjualan Pada Toko Sanwikarta" yang dipublikasikan pada JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika) Vol. 10, No. 2 bertujuan untuk memberikan rekomendasi tentang tren penjualan produk sembako dan menganalisis pola pembelian yang sering terjadi agar pemilik toko dapat mengembangkan strategi yang lebih tepat dalam penataan tata letak produk. Penelitian ini dilatarbelakangi oleh ketidakpahaman terhadap pola penjualan yang menyebabkan penurunan

efisiensi operasional dan ketidakpuasan pelanggan melalui pengelolaan stok yang tidak optimal, tata letak yang buruk, dan pengalaman belanja yang tidak memadai [18]. Metode yang digunakan adalah algoritma *FP-Growth* dengan bantuan aplikasi Rapid Miner untuk menganalisis tata letak barang yang akan dibeli secara bersamaan oleh pelanggan, dengan menerapkan proses pre-processing data untuk memastikan kualitas data yang digunakan dalam analisis. Hasil penelitian dengan menggunakan nilai minimum support 0.95 dan confidence 0.8 menunjukkan bahwa jika pelanggan membeli Saos Tomat, mereka juga akan membeli Mie Instan dan Sosis. Algoritma *FP-Growth* berhasil mengidentifikasi pola pembelian yang berulang, yang memungkinkan pemilik toko mengembangkan strategi yang lebih tepat untuk mengelola produk-produknya, menawarkan produk kepada pelanggan dengan lebih baik, dan pada akhirnya meningkatkan penjualan produk .

Penelitian yang dilakukan oleh Shen Keyi, et.al (2024) [19] ini bertujuan mengungkap hubungan kompleks antara kualitas udara dengan faktor meteorologi, morfologi perkotaan, dan sosial-ekonomi di 244 kota di China (2015-2019) menggunakan algoritma Apriori yang telah ditingkatkan. Latar belakang penelitian menunjukkan bahwa prediksi polusi udara sangat kompleks karena dipengaruhi oleh banyak faktor, sementara metode tradisional memiliki keterbatasan dalam kemudahan pemahaman hasil. Metode yang digunakan meliputi K-means clustering untuk mendiskritisasi data menjadi 5 cluster, kemudian menerapkan Improved Apriori Algorithm dengan penambahan metrik lift (threshold: support 1%, confidence 50%, lift >3) pada 1.033 titik data dari 10 prediktor (suhu, GDP, populasi, area hijau, dll.). Hasil penelitian menunjukkan peningkatan efisiensi komputasi yang drastis dengan pengurangan waktu sebesar 94,1% (dari 2.410 detik menjadi 141 detik) dan menghasilkan 361 aturan asosiasi yang kuat. Penelitian menemukan bahwa proporsi industri sekunder dalam GDP dan tekanan uap jenuh merupakan faktor paling berpengaruh terhadap kualitas udara. Polusi berat terjadi pada kombinasi pembangunan perkotaan dengan area hijau rendah dan suhu tinggi (confidence 52,4%).

Penelitian komparatif yang dilakukan oleh Rosadi dan Hasibuan (2024) berjudul "Comparison of Apriori and *FP-Growth* Algorithms in Analyzing Association Rules" yang dipublikasikan pada PIKSEL Vol. 12 (2) bertujuan untuk

membandingkan performa algoritma Apriori dan *FP-Growth* dalam analisis association rules dengan fokus pada efisiensi, akurasi, dan kecepatan pemrosesan data transaksi. Penelitian ini penting dilakukan karena pemilihan algoritma yang tepat sangat berpengaruh terhadap peningkatan efisiensi dan efektivitas pengambilan keputusan bisnis, khususnya di industri retail seperti toko sembako. Metodologi yang digunakan adalah pendekatan eksperimental empiris dengan melakukan perhitungan manual menggunakan 10 sampel data transaksi dan pengujian pada 38.765 data groceries yang diperoleh dari Kaggle menggunakan aplikasi Jupyter Notebook, dengan parameter yang seragam untuk kedua algoritma. Temuan penelitian menunjukkan bahwa algoritma *FP-Growth* terbukti jauh lebih efisien dengan waktu eksekusi 5.2757 detik dibandingkan Apriori yang memerlukan waktu 185.9585 detik, terutama dalam menangani dataset yang besar. Terdapat perbedaan pada hasil lift antara kedua algoritma, dimana Apriori sering menunjukkan hubungan negatif antar item ($\text{lift} < 1$), sedangkan *FP-Growth* memberikan interpretasi serupa namun dengan nilai lift yang sedikit berbeda, menunjukkan pengaruh yang berbeda dalam hubungan antar item pada data transaksi yang diuji [20].

Siddharta dan Trisnawarman sebelumnya juga menganalisis dataset cuaca dan kualitas udara Jakarta yang serupa, menggunakan uji Mann–Whitney U dan Korelasi Spearman. Temuannya menunjukkan kondisi panas-kering memperburuk kualitas udara, sementara kondisi lembap-hujan memperbaikinya. Namun, kedua uji tersebut hanya menganalisis hubungan dua variabel sekaligus, belum mengungkap kombinasi banyak kondisi yang saling terkait[21]. Oleh karena itu, penelitian ini menerapkan algoritma *FP-Growth* pada dataset yang sama untuk mengungkap pola kombinasi kondisi meteorologi dan polutan yang saling terkait terhadap kualitas udara di DKI Jakarta.

Tinjauan terhadap penelitian-penelitian tersebut menunjukkan bahwa algoritma *FP-Growth* telah banyak diterapkan di berbagai bidang. Oleh karena itu penelitian dengan judul “Penerapan *FP-Growth* untuk Mengidentifikasi Pola Asosiasi Polutan dan Faktor Meteorologi terhadap Kualitas Udara DKI Jakarta” yang menggunakan data ISPU dan meteorologi DKI Jakarta periode Januari 2024 hingga November 2025 yang diperoleh dari *website* Satu Data Jakarta. Algoritma

FP-Growth diterapkan dengan menguji kombinasi nilai minimum support 0.5, 0.6, dan 0.7 dan minimum confidence 0.5, 0.6, 0.7, 0.8, dan 0.9 untuk memperoleh aturan asosiasi yang paling kuat. Hasil penelitian diharapkan dapat memberikan pemahaman mengenai kondisi meteorologi dan parameter polutan yang berkaitan dengan kualitas udara di DKI Jakarta, sehingga dapat mendukung perumusan kebijakan pengendalian pencemaran udara yang lebih terarah.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut :

- 1) Bagaimana pola asosiasi antara kondisi meteorologi dan parameter polutan pada kualitas udara di DKI Jakarta menggunakan algoritma FP- Growth?
- 2) Apa saja aturan asosiasi terkuat yang terbentuk dari kombinasi kondisi meteorologi dan parameter polutan terhadap kualitas udara di DKI Jakarta berdasarkan nilai Support, Confidence dan Lift?

1.3 Tujuan Penelitian

Merujuk dari rumusan masalah diatas, adapun tujuan penelitian yang ingin dicapai peneliti dari penelitian ini adalah sebagai berikut:

1. Mengidentifikasi pola asosiasi antara kondisi meteorologi dan parameter polutan pada kualitas udara di DKI Jakarta menggunakan algoritma FP-Growth.
2. Mengidentifikasi aturan asosiasi terkuat yang terbentuk dari kombinasi kondisi meteorologi dan parameter polutan terhadap kualitas udara di DKI Jakarta berdasarkan nilai Support, Confidence, dan Lift.

1.4 Batasan Masalah

Dalam penelitian ini penting untuk menetapkan batasan yang jelas untuk memastikan hanya fokus pada isu-isu yang diteliti. Oleh karena itu, berikut batasan masalah yang diterapkan dalam penelitian ini :

1. Data yang digunakan dalam penelitian ini diambil dari website Satu Data Jakarta : <https://satudata.jakarta.go.id/home>.

2. Data yang digunakan adalah Indeks Standar Pencemaran Udara (ISPU) dan data meteorologi DKI Jakarta pada periode bulan Januari 2024 sampai dengan bulan November 2025.
3. Polutan yang dianalisis yaitu PM_{10} , $PM_{2.5}$, SO_2 , O_3 , NO_2 , CO , dan kategori.
4. Meteorologi yang dianalisis meliputi suhu udara, kelembaban udara, kecepatan angin, dan curah hujan.
5. Evaluasi aturan asosiasi menggunakan metrik support, confidence, dan lift.

1.5 Manfaat Penelitian

Pada manfaat penelitian ini dibagi menjadi dua, yaitu manfaat teoritis dan manfaat praktis.

1) Manfaat Teoritis

Penelitian ini diharapkan dapat memperluas pemahaman mengenai penerapan algoritma FP-Growth dalam menganalisis pola asosiasi antara kondisi meteorologi dan parameter polutan terhadap kualitas udara, serta memberikan kontribusi terhadap pengembangan metode analisis data mining dibidang lingkungan.

2) Manfaat Praktis

- a. Bagi penulis, penelitian ini dapat menambah wawasan dan pengalaman penulis dalam menerapkan algoritma FP-Growth untuk mengidentifikasi pola asosiasi pada data kualitas udara serta kemampuan dalam menganalisis data lingkungan.
- b. Bagi peneliti selanjutnya, penelitian ini dapat menjadi referensi atau dasar untuk pengembangan penelitian selanjutnya terkait analisis kualitas udara menggunakan metode data mining, khususnya dalam mengidentifikasi faktor-faktor yang berkaitan dengan kondisi kualitas udara.
- c. Bagi Pemerintah DKI Jakarta, khususnya Dinas Lingkungan Hidup (DLH) selaku pengelola Sistem Pemantauan Kualitas Udara (SPKU), hasil penelitian ini dapat menjadi pelengkap sistem

pelaporan yang saat ini hanya menampilkan kategori ISPU berdasarkan satu parameter pencemar dominan pada satu titik waktu. Pola dan aturan asosiasi yang dihasilkan memberikan gambaran mengenai kombinasi kondisi meteorologi dan parameter polutan yang secara konsisten muncul bersamaan dengan kategori kualitas udara SEDANG di masing-masing stasiun, sehingga dapat dimanfaatkan sebagai informasi pendukung dalam merumuskan kebijakan mitigasi yang lebih terarah, misalnya penentuan waktu atau lokasi prioritas untuk edukasi publik dan intervensi pengendalian emisi, serta sebagai bahan pertimbangan pengembangan sistem peringatan dini kualitas udara yang mempertimbangkan keterkaitan antar-parameter, bukan hanya satu parameter dominan seperti pada sistem yang berjalan saat ini.

1.6 Sistematika Penulisan

Sistematika penulisan proposal skripsi ini terdiri dari lima bab, yaitu sebagai berikut :

BAB I PENDAHULUAN

Bab ini memberikan gambaran umum mengenai penelitian yang akan dilakukan. Latar belakang menjelaskan alasan penting dan urgensi dari penelitian terhadap masalah yang akan diangkat, serta kedudukan masalah tersebut dalam konteks yang lebih luas. Selanjutnya, rumusan masalah menyajikan pertanyaan inti yang akan diteliti dalam bentuk kalimat tanya. Pada bagian batasan masalah, ruang lingkup penelitian dibatasi agar fokus dan terarah. Tujuan penelitian menjelaskan sasaran yang ingin dicapai, sedangkan manfaat penelitian menguraikan kontribusi teoretis dan praktis dari hasil penelitian. Terakhir, sistematika penulisan menggambarkan struktur keseluruhan skripsi secara ringkas.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan landasan teori dan konsep yang relevan dengan penelitian. Teori-teori yang digunakan sebagai kerangka berpikir diuraikan secara sistematis dan mendalam. Selain itu, disajikan pula penelitian-penelitian terdahulu yang memiliki keterkaitan dengan topik yang diteliti untuk menunjukkan posisi dan kebaruan penelitian ini di antara kajian-kajian yang telah ada sebelumnya. Pada bagian akhir bab, kerangka berpikir atau kerangka konseptual disusun untuk menggambarkan hubungan antara variabel-variabel yang diteliti.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode dan algoritma yang digunakan dalam penelitian, serta tahapan-tahapan yang dilakukan secara sistematis. Metode penelitian menguraikan pendekatan dan jenis penelitian yang dipilih, sedangkan algoritma yang digunakan dijelaskan secara rinci beserta alasan pemilihannya. Tahapan penelitian dimulai dari pengumpulan data yang menjelaskan sumber dan teknik pengambilan data. Selanjutnya, tahap preprocessing data menguraikan proses pembersihan, transformasi, dan persiapan data agar siap untuk dianalisis. Tahap pemodelan menjelaskan proses pembangunan model dengan algoritma yang telah dipilih, termasuk pengaturan parameter yang digunakan. Tahap evaluasi model menjelaskan metrik dan teknik yang digunakan untuk mengukur kinerja model yang telah dibangun. Terakhir, tahap implementasi menguraikan cara penerapan model yang telah dievaluasi untuk menyelesaikan permasalahan penelitian.

BAB IV HASIL DAN PEMABAHASAN

Penerapan metode yang telah dirancang serta evaluasi dari uji coba skenario yang dilakukan . Hasil pengumpulan data akan dianalisis dan dibahas.

BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan yang menjawab semua pertanyaan yang diajukan dalam bab sebelumnya.