

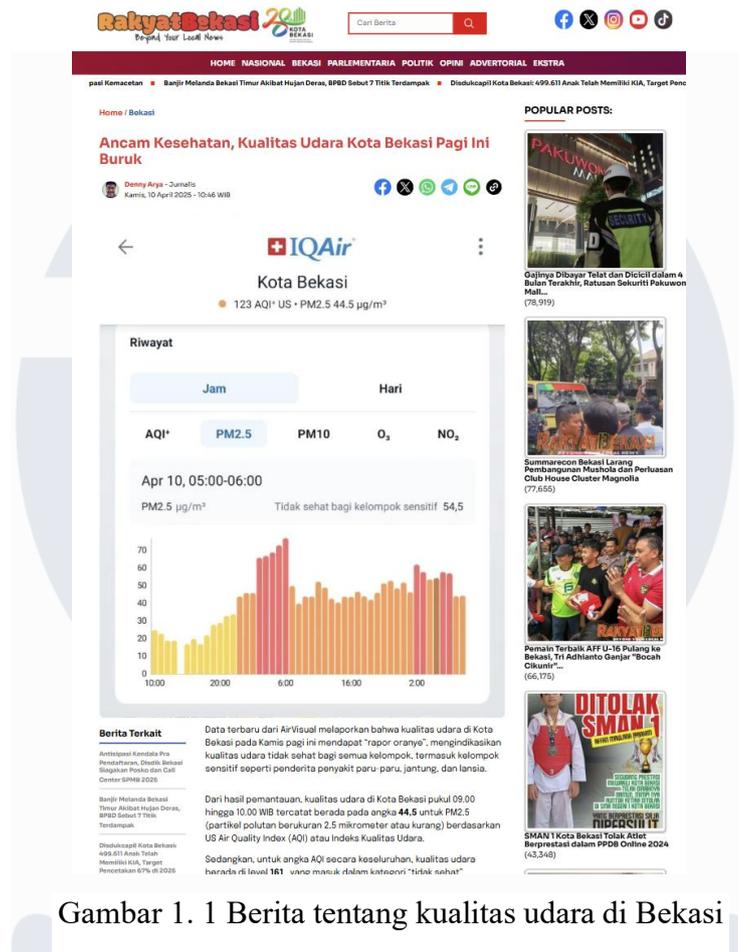
# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kota Bekasi selama beberapa tahun terakhir menunjukkan peningkatan jumlah penduduk yang cukup hal ini juga berdampak terhadap kualitas udara yang mengalami penurunan yang cukup signifikan. Kota ini menghadapi tantangan serius terkait pencemaran udara, yang sebagian besar disebabkan oleh meningkatnya emisi dari kendaraan bermotor, aktivitas industri, serta berbagai kegiatan manusia lainnya. Penurunan kualitas udara tersebut menimbulkan dampak yang nyata terhadap kesehatan masyarakat, menjadikannya sebagai isu lingkungan yang perlu mendapat perhatian lebih dari berbagai pihak [1]. Kualitas udara yang buruk dapat mengakibatkan berbagai masalah kesehatan serius, seperti Infeksi Saluran Pernapasan Akut (ISPA), asma, penyakit jantung, dan bahkan kematian [2] [22].





Gambar 1. 1 Berita tentang kualitas udara di Bekasi

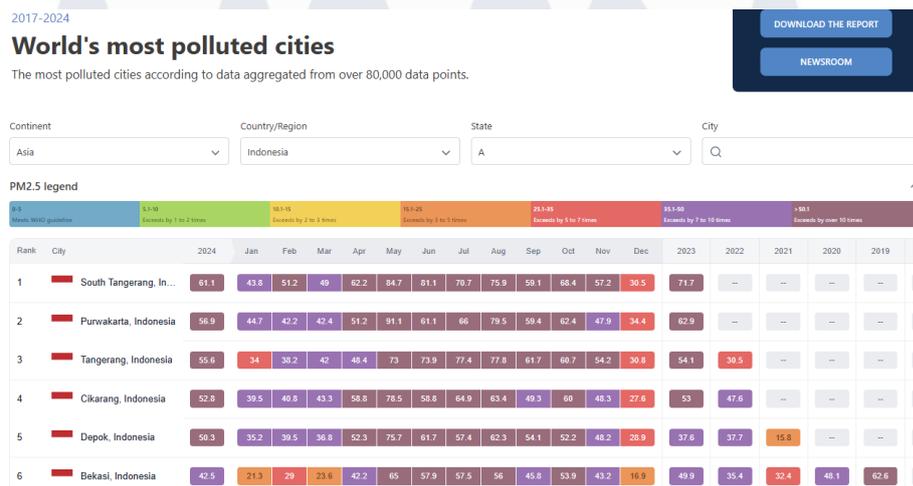
Sumber: [3]

Menurut berita dari situs rakyatbekasi.com yang ditulis oleh Denny Arya pada tanggal 10 April 2025, menyatakan bahwa kondisi kualitas udara di wilayah Bekasi pada pagi hari tanggal tersebut berada dalam situasi yang cukup mengkhawatirkan. Dalam laporan tersebut, kualitas udara mendapatkan penilaian dalam kategori “rapor oranye”, yang secara umum menandakan bahwa udara berada dalam level yang tidak sehat, bahkan berisiko bagi seluruh lapisan masyarakat tanpa terkecuali, termasuk individu yang sebelumnya tidak memiliki gangguan pernapasan. Lebih lanjut, tingkat konsentrasi partikel PM2.5 sebagai

salah satu tolak ukur dalam pengukuran polusi udara tercatat berada pada angka 161. Angka ini secara jelas menempatkan kualitas udara Bekasi ke dalam klasifikasi “tidak sehat” menurut standar indeks kualitas udara, yang dapat berdampak negatif terhadap kesehatan masyarakat dalam jangka pendek maupun panjang jika kondisi tersebut berlangsung secara terus-menerus.

Salah satu indikator yang sering dipakai untuk menilai tingkat kualitas udara merupakan IQAir, yang merupakan alat untuk mengukur kualitas udara secara global. IQAir mencakup berbagai parameter, seperti Particulate Matter (PM), nitrogen oksida (NO<sub>x</sub>), karbon monoksida (CO), sulfur oksida (SO<sub>x</sub>), yang digunakan untuk menilai tingkat pencemaran udara dalam suatu wilayah [4]. Data IQAir digunakan untuk memantau kualitas udara secara global.

Berdasarkan data dari pemantau kualitas udara IQAir menyatakan bahwa Indonesia berada pada peringkat 3 sebagai kota paling polusi di dunia [4] yang sebagian besar berasal dari emisi kendaraan dan hasil pembakaran batu bara. Hal tersebut berdampak pada masyarakat yang memicu untuk mendapatkan penyakit pernapasan, penyakit jantung, serta kematian.



Gambar 1. 2 Grafik Rangking kota paling berpolusi di Indonesia

Sumber: [16]

Dalam gambar 1.2 ini menunjukkan daftar kota paling berpolusi di dunia, dengan fokus pada wilayah Asia, khususnya Indonesia, dalam periode 2017-2024. Data ini didasarkan pada pengukuran PM2.5, yang merupakan partikel-partikel halus di udara dengan ukuran kurang dari 2,5 mikrometer. Partikel ini dianggap berbahaya karena kemampuannya menembus saluran pernapasan dan berpotensi menimbulkan berbagai masalah kesehatan. Berdasarkan data pemantauan kualitas udara dari situs IQAir, beberapa wilayah di Bekasi pada tahun 2022 mengalami indeks kualitas udara (Air Quality Index/AQI) yang mencapai angka 46.1, yang menurut standar internasional termasuk dalam kategori “tidak sehat”. Angka ini menunjukkan bahwa udara di wilayah tersebut mengandung partikel dan zat polutan yang dapat membahayakan kesehatan, tidak hanya bagi kelompok rentan seperti anak-anak dan lansia, tetapi juga bagi masyarakat umum [4].

Penelitian ini bertujuan untuk mengambil data IQAir terhadap kualitas udara yang tercermin dalam indeks pencemaran udara. Hasil analisis ini diharapkan dapat membantu pengambilan kebijakan yang lebih efektif dalam upaya meningkatkan kualitas udara dan perlindungan kesehatan masyarakat. Analisis sentimen juga dapat membantu dalam mengidentifikasi faktor-faktor yang memengaruhi klasifikasi udara. Penelitian sebelumnya menunjukkan bahwa algoritma Support Vector Machine (SVM), Naive Bayes, dan Random Forest banyak digunakan dalam berbagai kasus klasifikasi, termasuk klasifikasi data berbahaya, data lingkungan, dan situs phishing. Penelitian oleh Abdullah Batuhan Yilmaz, Yavuz Selim Taspinar, dan Murat Koklu (2024) menunjukkan bahwa algoritma Naive Bayes mampu mencapai akurasi sebesar 92,4% dalam mengklasifikasikan aplikasi Android berbahaya, sementara SVM memiliki akurasi sebesar 90,9% [5]. Hasil ini menunjukkan bahwa Naive Bayes memiliki keunggulan dalam menangani data berbasis teks dengan karakteristik tertentu.

Penelitian lain yang dilakukan oleh Bashir Olasunkanmi Ayinde, Muhammed Rabiul Musa, dan Abdul-Afeez Olakunle Ayinde (2024) menggunakan algoritma Random Forest dan XGBoost dalam memodelkan konsentrasi PM2.5 berdasarkan citra satelit Landsat 8 [6]. Penelitian tersebut menggabungkan variabel lingkungan seperti penggunaan lahan dan data permukaan untuk melatih model dalam memprediksi kadar polusi udara. Di antara model yang diuji, XGBoost menunjukkan performa terbaik dalam hal akurasi dan kestabilan prediksi, khususnya pada musim kemarau yang cenderung bebas dari gangguan atmosfer. [6]. Peneliti menggunakan data permukaan, informasi lahan, dan variabel lingkungan lainnya dari citra satelit untuk melatih model dalam memprediksi kadar PM2.5 di berbagai musim. Hasilnya menunjukkan bahwa model berbasis *machine learning*, terutama Random Forest, mampu menghasilkan estimasi PM2.5 yang cukup akurat, dengan kinerja terbaik terjadi pada musim kemarau karena citra lebih bersih dari gangguan atmosfer seperti awan atau hujan. Di antara model yang diuji, XGBoost menunjukkan performa terbaik dalam hal akurasi dan stabilitas prediksi, menjadikannya metode yang efektif untuk pemantauan kualitas udara di daerah dengan keterbatasan stasiun pemantauan konvensional.

Selain itu, penelitian oleh Nabila Bianca Putri dan Arie Wahyu Wijayanto (2022) membandingkan algoritma Naive Bayes, Random Forest, dan SVM dalam klasifikasi website phishing. Hasilnya menunjukkan bahwa Random Forest memberikan akurasi tertinggi yaitu 90,77%, diikuti oleh SVM dengan akurasi 86,25% [7]. Hal ini memperkuat temuan bahwa Random Forest memiliki keunggulan dalam menangani data yang bersifat kompleks dan bervariasi. Selain itu, penelitian oleh Nabila Bianca Putri dan Arie Wahyu Wijayanto (2022) membandingkan algoritma Naive Bayes, Random Forest, dan SVM dalam klasifikasi website phishing. Hasilnya menunjukkan bahwa Random Forest memberikan akurasi tertinggi yaitu 90,77%, diikuti oleh SVM dengan akurasi

86,25% [7]. Hal ini memperkuat temuan bahwa Random Forest memiliki keunggulan dalam menangani data yang bersifat kompleks dan bervariasi.

Penelitian ini bertujuan untuk melakukan perbandingan terhadap kinerja tiga algoritma klasifikasi dalam menganalisis data kualitas udara di Kota Bekasi, dengan memanfaatkan data historis Air Quality Index (AQI) selama periode 2021 hingga 2025. Hasil yang diperoleh diharapkan mampu memberikan kontribusi dalam pengembangan teori di bidang *data mining*, sekaligus menghasilkan solusi praktis berupa sistem klasifikasi kualitas udara yang dapat dimanfaatkan oleh pemerintah kota serta masyarakat dalam mendukung perbaikan lingkungan dan kesehatan masyarakat.

## **1.2 Rumusan Masalah**

Dari latar belakang yang telah dijelaskan, dapat dirumuskan beberapa permasalahan inti yang akan ditelusuri dalam penelitian ini, antara lain:

1. Bagaimana algoritma Support Vector Machine, XGBoost dan Random Forest digunakan dalam membangun model klasifikasi kualitas udara berdasarkan data AQI?
2. Parameter pencemar udara apa saja (seperti PM10, SO<sub>2</sub>, CO, NO<sub>2</sub>, dan O<sub>3</sub>) yang paling berpengaruh dalam proses klasifikasi kualitas udara di Kota Bekasi?
3. Di antara algoritma SVM, XGBoost dan Random Forest, manakah yang memberikan hasil klasifikasi paling akurat berdasarkan metrik evaluasi seperti akurasi, presisi, recall, dan f1-score?

## **1.3 Batasan Masalah**

Mengacu pada rumusan masalah yang telah dijabarkan sebelumnya, perlu dilakukan pembatasan ruang lingkup pembahasan agar arah penelitian tetap terfokus dan tidak melebar ke isu-isu di luar topik utama, antara lain:

1. Data yang digunakan merupakan data sekunder yang diperoleh dari situs

Air Quality Index (AQI) di kota Bekasi dan dibatasi pada tanggal 1 Januari 2021 hingga 4 Mei 2025. Data mencakup lima parameter utama polutan, yaitu PM2.5, PM10, SO<sub>2</sub>, CO, O<sub>3</sub>, dan NO<sub>2</sub>, dan tidak mencakup parameter tambahan seperti suhu dan kelembapan.

2. Sumber data hanya diambil dari situs Air Quality Index (AQI), dengan data sekunder yang tersedia secara publik dari situs tersebut.
3. Penelitian ini tidak mengevaluasi atau membahas dampak kesehatan secara langsung dari klasifikasi kualitas udara yang dihasilkan, melainkan hanya berfokus pada evaluasi performa model algoritma klasifikasi berdasarkan metrik akurasi, presisi, recall, dan f1-score.

#### **1.4 Tujuan dan Manfaat Penelitian**

##### **1.4.1 Tujuan Penelitian**

Tujuan utama dari penelitian ini adalah melakukan analisis data kategori udara selama 4 tahun terakhir:

1. Menghasilkan klasifikasi pada Support Vector Machine, XGBoost, dan Random Forest.
2. Mengukur performa 3 algoritma tersebut dari akurasi dalam klasifikasi kualitas udara.
3. Membandingkan kinerja algoritma SVM, XGBoost dan Random Forest untuk mendukung pemilihan metode terbaik dalam klasifikasi berbasis AQI.

##### **1.4.2 Manfaat Penelitian**

Berdasarkan rumusan masalah yang telah dirumuskan tersebut, peneliti memiliki manfaat dalam penelitian ini, yaitu:

- **Manfaat Teoritis:**

Berperan dalam meningkatkan dan mengembangkan pengetahuan di bidang *data mining* dan *machine learning*, khususnya dalam penerapan algoritma klasifikasi untuk analisis

kualitas udara.

- **Manfaat Praktis:**

1. Memberikan informasi kepada pihak terkait mengenai kategori kualitas udara di Kota Bekasi selama periode 2021–2025.
2. Menyediakan referensi perbandingan kinerja algoritma klasifikasi untuk mendukung pengambilan keputusan dalam pemantauan dan pengelolaan kualitas udara.

## 1.5 Sistematika Penulisan

### BAB 1 PENDAHULUAN

Bab ini menjelaskan alasan dan konteks pentingnya penelitian ini dilakukan. Rumusan masalah menguraikan pertanyaan-pertanyaan utama yang ingin dijawab melalui penelitian. Batasan masalah menetapkan ruang lingkup agar penelitian tetap fokus dan terarah. Tujuan penelitian menjelaskan hasil yang ingin dicapai, sedangkan manfaat penelitian menggambarkan kontribusi dan kegunaan dari hasil penelitian tersebut. Terakhir, sistematika penulisan memberikan gambaran struktur dan susunan isi laporan skripsi secara keseluruhan.

### BAB 2 LANDASAN TEORI

Pada bab ini, berisikan tentang teori-teori yang digunakan dalam penelitian skripsi dan penjelasan dari *framework* dan algoritma serta *tools* yang digunakan. Teori-teori yang dimaksud merupakan teori mengenai udara, kualitas udara, IQAir, *framework* CRISP-DM, algoritma Support Vector Machine (SVM), XGBoost dan Random Forest serta *tools* yang digunakan, yaitu Jupyter Notebook, Python, dan Anaconda sebagai lingkungan pengembangan dan pemrograman dalam penelitian ini.

### BAB 3 METODOLOGI PENELITIAN

Pada bab ini, berisikan tentang proses penelitian yang dilakukan secara bertahap. Terdapat berbagai diagram dan gambar yang disertakan dengan penjelasan mengenai alur penelitian dan teknik pengambilan data yang diterapkan dalam penelitian ini.

#### BAB 4 ANALISIS DAN HASIL PENELITIAN

Bab 4 menyajikan hasil yang diperoleh dari seluruh proses pengumpulan data yang telah dilakukan. Data tersebut kemudian diolah dan dianalisis dengan menggunakan metode yang relevan agar dapat menjawab pertanyaan-pertanyaan penelitian yang telah dirumuskan. Dalam bab ini, hasil analisis disajikan secara terperinci untuk memberikan gambaran yang komprehensif mengenai temuan penelitian serta hubungannya dengan masalah yang sedang dikaji.

#### BAB 5 SIMPULAN DAN SARAN

Bab 5 berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran-saran yang dapat diberikan berdasarkan temuan tersebut. Bab ini merupakan bagian akhir dari laporan penelitian dan bertujuan untuk merangkum inti dari penelitian secara keseluruhan. Menjelaskan secara ringkas mengenai capaian dari penelitian serta mengarahkan pembaca pada implikasi dari hasil penelitian tersebut.

