

BAB I PENDAHULUAN

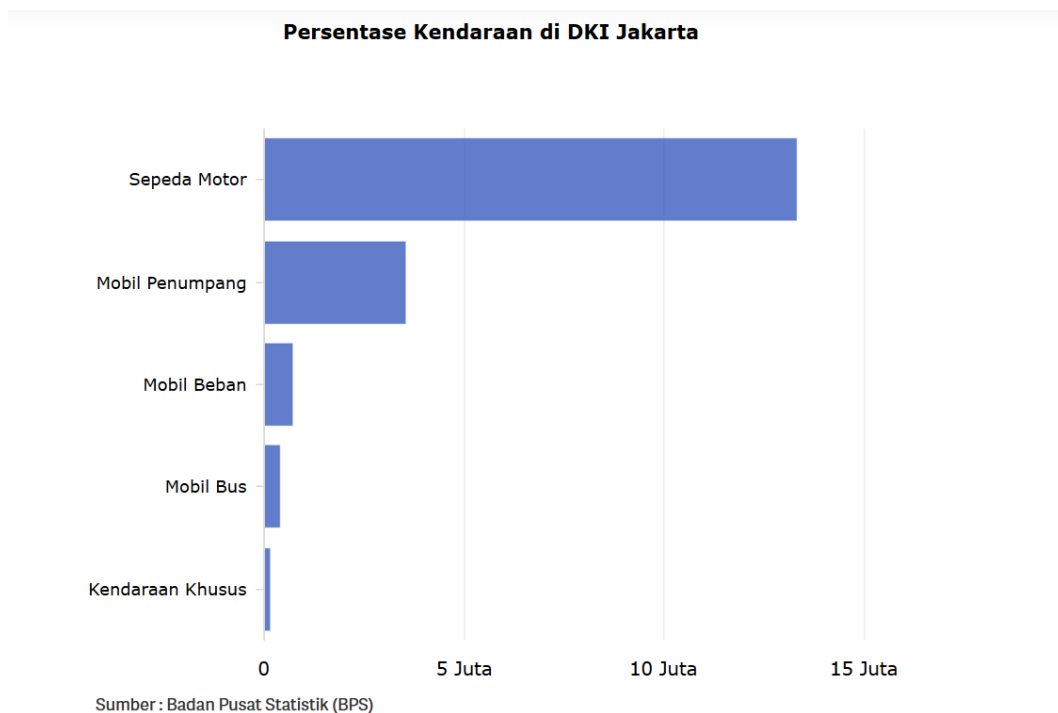
I.1. Latar Belakang

Pencemaran udara merupakan salah satu masalah yang cukup mengkhawatirkan yang dikenal dunia sejak ratusan tahun lalu (Mukono, 2011). Selama lima puluh tahun terakhir, kualitas udara global telah menunjukkan penurunan, efek langsung dari aktivitas manusia seperti pembakaran biomassa, operasi industri dan emisi kendaraan (Adame, Notario, Villanueva, & Albaladejo, 2012).

Dampak terhadap kesehatan yang disebabkan oleh pencemaran udara akan terakumulasi dari hari ke hari. Pemaparan dalam jangka waktu lama akan berakibat pada berbagai gangguan kesehatan, seperti *bronchitis*, *emphysema*, dan kanker paru-paru. Dampak kesehatan yang diakibatkan oleh pencemaran udara berbeda-beda antar individu. Populasi yang paling rentan adalah kelompok individu berusia lanjut dan balita (Tarumingkeng, Coto, & Hardjanto, 2004).

Penyebab pencemaran udara bisa berasal dari kegiatan rumah tangga, emisi kendaraan bermotor dan kegiatan industri (Mukono, 2011). Selain itu, beberapa penelitian (Zhang & Smith, 2007; Harinath & Murthy, 2012; Laumbach & Kipen, 2012; Bergstra, Brunekreef, & Burdorf, 2018), telah menunjukkan hubungan antara partikel dan pernapasan akut dan penyakit kardiovaskular, di antara banyak lainnya. Karena itu, banyak pekerjaan penelitian telah dilakukan terkait dengan polusi udara dan pemantauan kualitas udara, identifikasi sumber, jalur transportasi polutan jarak jauh dan pengembangan dan implementasi strategi pengendalian dan mitigasi yang efektif.

Menurut Ridhoi (2020), Polusi Udara di Jakarta Menurun Sebagaimana dikatakan Anies Baswedan, sebetulnya penduduk Jakarta telah membatasi pergerakan di luar rumah sejak beberapa minggu lalu. Membuat kondisi jalanan di Jakarta menjadi lebih sepi dari sebelum korona merebak. Jakarta sebelumnya memang kota sibuk.



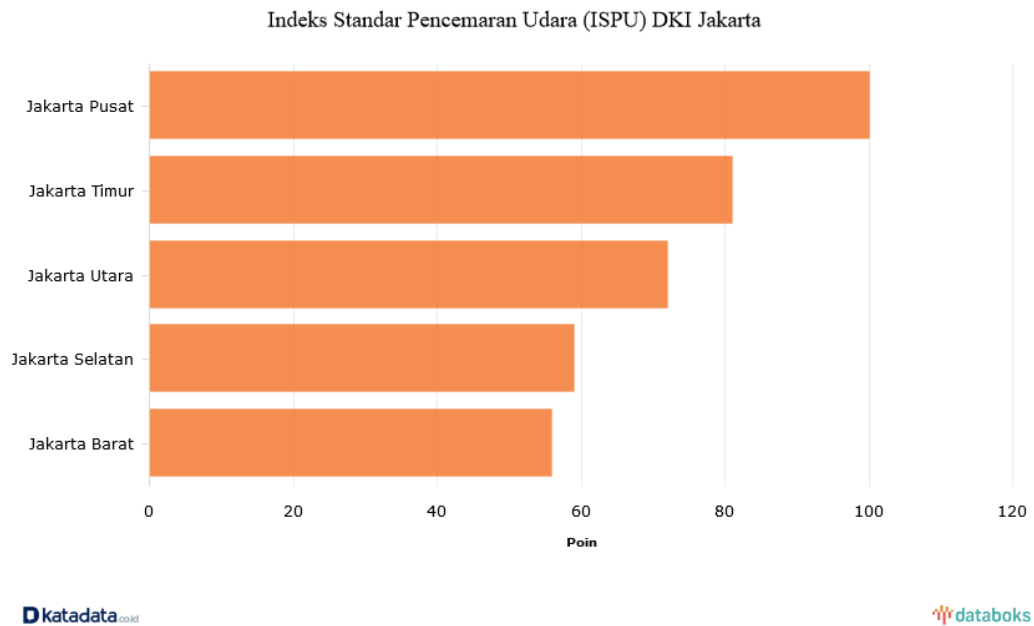
Gambar I 1 Presentase Kendaraan di DKI Jakarta

(Sumber: <https://katadata.co.id>)

Pada gambar I.1 dapat dilihat presentase kendaraan di DKI Jakarta paling tinggi yaitu pada Sepeda motor sebanyak 13,5 Juta pengguna, kedua mobil penumpang dan yang ketiga mobil beban. Maka dari itu peneliti melakukan penelitian untuk menentukan clustering area pada Provinsi DKI Jakarta.

Data Kementerian Lingkungan Hidup dan Kehutanan (KLHK) menyatakan, polusi udara di Provinsi DKI Jakarta masuk dalam kategori sedang menurut Indeks Standar Pencemaran Udara (ISPU) (Jayani, 2019). Kategori sedang berskala 51-101 yang menunjukkan tingkat kualitas udara yang tidak memberikan efek bagi kesehatan manusia atau hewan, tetapi berpengaruh pada tumbuhan yang sensitif dan nilai estetika. Meskipun tergolong sedang, nilai ISPU ini perlu diwaspadai karena terdapat wilayah yang mendekati atau di ambang batas menuju kategori tidak sehat.

KLHK memiliki sejumlah stasiun pemantau kualitas udara di DKI Jakarta. Di Jakarta Timur terdapat di Jalan Raya Pondok Gede, di Jakarta Utara ada di Kantor Kelurahan Kelapa Gading, dan di Jakarta Selatan di Taman Pendidikan Dinas Pertamanan. Di Jakarta Barat terdapat di Taman Perumahan Kebon Jeruk sedangkan di Jakarta Pusat ada beberapa lokasi, termasuk di dekat Pos Polisi Bundaran HI (Jayani, 2019).



Gambar I 2 Indeks Standar Pencemaran Udara (ISPU) DKI Jakarta

(Sumber : <https://databoks.katadata.co.id>)

Pada gambar I.2 dapat dilihat Jakarta Pusat memiliki nilai ISPU mengkhawatirkan, yaitu senilai 100. Nilai ISPU tertinggi selanjutnya terdapat di Kota Jakarta Timur (81) dan Jakarta Utara (72). Sementara Jakarta Barat yang memiliki nilai 56 dan Jakarta Selatan yang memiliki nilai 59 masih dalam batas aman.

Dalam penelitian ini, dilakukan *clustering* Indeks Standar Pencemaran Udara (ISPU) di Jakarta karena agar kita mengetahui *cluster* mana saja yang pada wilayah/stasiun tersebut terdapat kategori yang tidak sehat meskipun Indeks Standar Pencemaran Udara (ISPU) berbeda-beda dari segi daerah maupun iklim (Renaldi, 2017). Maka dari itu, analisis *cluster* telah menjadi alat yang efektif untuk analisis polutan udara. Analisis *cluster*, atau lebih umum disebut sebagai "*clustering*", adalah teknik yang digunakan untuk pengelompokan pengamatan yang serupa, titik data atau vektor fitur berdasarkan karakteristik yang sama (Jain, Murty, & Flynn, 1999). Menurut Kaufman & Rousseeuw (2009), itu adalah "seni menemukan kelompok dalam data". Secara umum, tujuan analisis *cluster* adalah untuk mengidentifikasi kelompok-kelompok objek yang serupa, di mana objek-objek dalam satu kluster lebih mirip satu sama lain daripada objek-objek di *cluster* yang berbeda.

Clustering dapat digunakan untuk identifikasi pola dan distribusi yang menarik dan menghasilkan wawasan yang mungkin ke dalam struktur data yang mendasarinya (Halkidi, Batistakis, & Vazirgiannis, 2001). Oleh karena itu, analisis *cluster* adalah teknik yang berguna untuk menemukan dan mengekstraksi informasi yang mungkin sebelumnya tidak diperhatikan. Analisis *cluster* diusulkan pada awal tahun 1930, namun penerapannya baru populer pada tahun 1960-an.

Clustering merupakan masalah pengelompokan data yang ditetapkan ke dalam beberapa kelompok sedemikian rupa sehingga, definisi yang kurang dari "kesamaan," item yang sama berada di kelompok yang sama dan berbeda dalam kelompok yang berbeda (Guha & Mishra, 2016). *Clustering* K-Means melibatkan pengambilan model yang sesuai dengan data historis dan menggunakannya untuk meng*cluster*. Beberapa metode untuk menyesuaikan model adalah K-*Means*, K-*Modes*, K-*Medoids*, dan *Fuzzy c- Means* (Suyanto, 2017). Pada penelitian kali ini digunakan K-*Means* untuk membuat model *Clustering* Indeks Standar Pencemaran Udara. K-*Means* merupakan pengelompokan data yang umum digunakan untuk melakukan tugas belajar tanpa pengawasan (Ding & He, 2004).

Berdasarkan Rodiyansyah (2017), dilakukan *K-Means* dengan *Fuzzy c-Means* pada analisis data polusi udara di kota X. Didapatkan hasil rata-rata standar deviasi pada hasil *clustering fuzzy c-means* lebih kecil daripada rata-rata standar deviasi pada hasil *clustering K-Means*. Dengan hasil nilai cluster bahwa parameter natrium dioksida, non metal hydro carbon dan natrium oksida memiliki pengaruh yang signifikan. Pada penelitian Yazid & Affandes (2018) digunakan *K-Means* untuk membuat model *clustering* data pencemaran udara di Provinsi Riau dengan model terbaik yang didapatkan yaitu *K-Means* terdapat 16 bulan dari 84 bulan dari tahun 2009 – 2015 yang memiliki kadar polutan PM10 yang termasuk ke dalam kategori sedang hingga sangat tidak sehat yang digunakan untuk melakukan *clustering* merupakan jumlah cluster setiap bulannya.

Pada Standar Indeks Pencemaran Udara (ISPU) kenapa perlu di *clustering*, karena di perlukan *clustering* berdasarkan parameter yang mengandung zat-zat kimia berbahaya yang dihasilkan dari berbagai aktivitas pada faktor internal dan eksternal penyebab pencemaran udara. Hasil dari *clustering* ini dapat digunakan sebagai informasi untuk membantu pemerintah daerah khususnya dalam mengambil kebijakan dan menentukan langkah yang tepat dalam mencegah dan mengatasi pencemaran udara.

Berdasarkan penelitian sebelumnya bahwa *K-Means* lebih bagus dalam menggunakan algoritma pada case ini, karena karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam cluster yang lain.

Berdasarkan hal tersebut maka akan dibuatlah sebuah rancangan data mining yang akan membagi data polutan yang sangat besar ke dalam beberapa cluster. Pengelompokan data polutan udara dapat dilakukan dengan menggunakan 0-2 analisis cluster dengan memanfaatkan metode *K-means*.

K-means merupakan metode *clustering* yang paling sederhana dan umum.

K-means mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien. Alasan pemilihan algoritma K-Means, data sebanyak 5382 data dengan kriteria awalnya *object* menjadi *int*.

Algoritma *K-Means* dipilih karena jauh lebih baik dari model yang lainnya. Dalam melakukan penelitian menggunakan model algoritma *K-Means* untuk melakukan *clustering* data Indeks Standar Pencemaran Udara (ISPU) pada Provinsi DKI Jakarta. Hasil dari penelitian ini diharapkan dapat membantu tiap *cluster* Indeks Pencemaran Udara untuk setiap tahunnya pengaruh terhadap tiap parameter pencemaran udara serta agar masyarakat lebih waspada, untuk mencegah efek negative dari pencemaran udara seperti ispa dan gangguan pernapasan lainnya bahkan dapat menyebabkan kematian.

I.2. Perumusan Masalah

Rumusan masalah yang ada pada penelitian ini adalah sebagai berikut.

1. Bagaimana penerapan algoritma *K-Means clustering* terhadap data Indeks Standar Pencemaran Udara (ISPU) Provinsi DKI Jakarta ?
2. Bagaimana hasil penerapan algoritma *K-Means clustering* terhadap data Indeks Standar Pencemaran Udara (ISPU) Provinsi DKI Jakarta ?

I.3. Tujuan Penelitian

Adapun Tujuan penelitian ini adalah, sebagai berikut.

1. Menerapkan algoritma *K-Means clustering* untuk mengelompokkan parameter mana yang tertinggi atau berbahaya berdasarkan analisis data parameter udara terhadap data Indeks Standar Pencemaran Udara (ISPU) Provinsi DKI Jakarta.
2. Mengetahui hasil penerapan algoritma *K-Means clustering* dan memberikan informasi berdasarkan tiap cluster yang diperoleh dari hasil clustering dengan algoritma *K-Means* terhadap data Indeks Standar Pencemaran Udara (ISPU) Provinsi DKI Jakarta.

I.4. Batasan Penelitian

Batasan dari penelitian ini adalah sebagai berikut:

1. Penelitian hanya dilakukan pada Indeks Standar Pencemaran Udara (ISPU) di Provinsi DKI Jakarta dari tahun 2017-2019 yang diperoleh dari situs Jakarta Open Data dan Dinas Lingkungan Hidup Provinsi DKI Jakarta;
2. Parameter pencemaran udara yang digunakan hanya zat Carbon Monoxide (CO), Partikulat (PM10), Sulfur Dioksida (SO₂), Nitrogen Dioksida (NO₂), dan Ozon (O₃);
3. Metode yang digunakan adalah *Clustering* data pada Metode *K-Means*;
4. Bahasa Pemrograman yang akan digunakan adalah *Python*;
5. *Clustering* hanya dilakukan untuk bulan januari tahun 2017 sampai dengan bulan Desember tahun 2019.

I.5. Manfaat Penelitian

Manfaat dari pengerjaan tugas akhir adalah:

1. hasil penelitian ini diharapkan bisa menjadi acuan atau saran bagi penelitian tentang *Clustering* data Indeks Standar Pencemaran Udara (ISPU) Provinsi DKI Jakarta pada tahun 2017-2019 sehingga informasi ini dapat digunakan untuk mengetahui tingkat kualitas udara dari tiap pertahun agar masyarakat lebih menjaga lingkungan agar kualitas udara menjadi lebih baik;
2. Metoda *K-Means* pakai diharapkan dapat diterapkan untuk *Clustering* ISPU di masa depan secara otomatis;
3. penelitian ini diharapkan dapat membantu pembelajaran penerapan suatu metoda *machine learning* di Fakultas Rekayasa Industri Prodi Sistem Informasi .

I.6. Metodologi Penelitian

Metodologi dalam penyelesaian tugas akhir ini antara lain:

1. Studi literatur, pengumpulan referensi berupa artikel ilmiah yang diambil

dari beberapa jurnal maupun dari situs internet;

2. Studi kasus yang telah dilakukan dari hasil pengumpulan jurnal yang diamati;
3. Analisis data yang telah didapatkan dengan melihat data tersebut dapat diolah menggunakan metode clustering atau tidak serta melihat hasil yang didapatkan;
4. Penyusunan laporan Tugas Akhir akan menjelaskan penelitian yang dilakukan sampai dengan hasil yang diperoleh.

I.7. Sistematika Penelitian

Penelitian ini di uraikan ke dalam beberapa bab melalui sistematika penulisan sebagai berikut:

BAB I Pendahuluan

Bab ini berisi penjelasan mengenai latar belakang penelitian, perumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, dan sistematika penulisan.

BAB II Landasan Teori

Bab ini berisi teori-teori yang relevan dengan permasalahan yang diteliti sehingga dapat digunakan untuk menunjang pelaksanaan penelitian.

BAB III Metodologi Penelitian

Bab ini berisi penjelasan mengenai langkah-langkah penelitian secara rinci meliputi model konseptual dan metode yang digunakan pada penelitian.

BAB IV Analisis dan Perancangan

Bab ini berisi tentang analisis perancangan sistem dengan melakukan analisis kebutuhan pengguna.

BAB V Pengujian dan Analisis Hasil Pengujian

Bab ini berisi tentang pengujian dari hasil perancangan yang telah dibuat sebelumnya, kemudian dilakukan analisis hasil pengujian sehingga mendapat keluaran berupa perbaikan *clustering* data.

BAB VI Kesimpulan dan Saran

Bab ini menjelaskan kesimpulan dari sistem yang telah dibangun sesuai dengan rancangan dan implementasi yang telah dilakukan, serta saran yang diberikan pada saat melakukan pengujian dan melakukan operasionalisasi variabel penelitian, menyusun kuesioner penelitian, merancang pengumpulan dan pengolahan data, melakukan uji instrumen, merancang analisis pengolahan data.