

BAB I

PENDAHULUAN

1.1 Latar Belakang

Jakarta sebagai ibu kota negara Republik Indonesia menjadi pusat berbagai kegiatan diiringi dengan laju urbanisasi yang pesat. Semakin bertambahnya pengguna kendaraan pribadi menjadi penyumbang utama buruknya kualitas udara di ibu kota.

Menurut laporan yang dikeluarkan oleh *Greenpeace* India pada tahun 2022 mengenai ketimpangan paparan polusi udara di beberapa negara, diperkirakan setiap harinya sekitar 93% warga Jakarta harus menghirup udara berbahaya dengan konsentrasi polutan lima kali lebih besar dari standar batas aman yang direkomendasikan WHO (Organisasi Kesehatan Dunia). Konsentrasi rata-rata tahunan $PM_{2.5}$ (polusi dalam bentuk partikel yang berdiameter 2,5 mikro atau kurang) di Jakarta mencapai 25 mikrogram per meter³, lima kali lebih besar dari batas aman yang direkomendasikan WHO yaitu 5 mikrogram per meter³ (Greenpeace India, 2022).

Dengan tingginya nilai konsentrasi polutan $PM_{2.5}$ akan memberikan efek kesehatan negatif secara langsung pada kelompok masyarakat rentan, seperti orang dengan masalah pernafasan, anak-anak, ibu hamil, dan orang lanjut usia. Bahkan bila terpapar dalam jangka waktu yang lama, konsentrasi $PM_{2.5}$ yang tinggi dapat merugikan kesehatan orang yang dalam kondisi baik.

Dalam melaksanakan pengendalian kualitas udara, salah satu langkah yang dapat dilakukan adalah melalui informasi terkini mengenai kualitas udara di sekitar. Pentingnya kesadaran akan kualitas udara di sekitar memungkinkan setiap individu untuk membuat keputusan yang tepat untuk membatasi aktifitas diluar ruangan dan mengurangi paparan polusi, menyadari dampak polusi terhadap lingkungan, dan tanggung jawab bersama untuk menciptakan dunia yang lebih bersih dan sehat.

Sesuai dengan peraturan Kementerian Lingkungan Hidup dan Kehutanan, pemerintahan wilayah setempat setiap harinya menyediakan informasi publik mengenai hasil penentuan kategori Indeks Standar Pencemar Udara (ISPU). ISPU menggambarkan kondisi mutu udara ambien di lokasi tertentu, yang didasarkan pada dampak terhadap kesehatan manusia (Peraturan Menteri Lingkungan Hidup dan Kehutanan nomor 14 tahun 2020 tentang Indeks Standar Pencemar Udara).

Melalui 5 stasiun pemantauan kualitas udara yang ada di wilayah DKI Jakarta, dilakukan pemantauan pada 6 parameter yakni partikulat (PM_{10} dan $PM_{2.5}$), nitrogen dioksida (NO_2), sulfur dioksida (SO_2), karbon monoksida (CO), ozon (O_3). Dari hasil pemantauan parameter tersebut kemudian dilakukan konversi menjadi nilai ISPU untuk selanjutnya diklasifikasikan kedalam 5 kategori yaitu baik, sedang, tidak sehat, sangat tidak sehat, dan berbahaya.

Berdasarkan informasi pemantauan konsentrasi polutan dan kategori ISPU yang telah dilakukan sampai saat ini, beberapa peneliti menemukan peluang untuk memanfaatkan metode *machine learning* dalam klasifikasi indeks kualitas udara. Menurut Liang, et al., 2022, *machine learning* melibatkan metode komputasi yang

belajar dari data kompleks untuk membangun berbagai model untuk prediksi, klasifikasi, dan evaluasi (Liang, Y.-C., et al., 2020). Klasifikasi termasuk kedalam *supervised learning* yang membangun model berdasarkan data dan variabel kelas. Model klasifikasi memungkinkan untuk memahami karakteristik data dari setiap kelas dan dapat digunakan untuk mengidentifikasi kelas data baru (Cheng & Tsai, 2022).

Tlais, et al. pada tahun 2020 melakukan penelitian mengenai monitoring dan klasifikasi kualitas udara dengan menerapkan *machine learning* pada dataset polusi udara di Australia. Hasil penelitian tersebut menunjukkan bahwa metode *Multilayer Perceptron* (MLP) menghasilkan nilai akurasi tertinggi dibandingkan dengan 3 metode *machine learning* lainnya, akurasi MLP mencapai 99%. Menurut Tlais, *machine learning* merupakan pendekatan yang menjanjikan untuk menyelesaikan kasus prediksi kualitas udara yang akurat (Tlais, et al., 2020).

Shaziayani, et al. pada tahun 2022 juga melakukan penelitian mengenai klasifikasi prediksi konsentrasi PM₁₀ menggunakan *machine learning* berbasis *tree*. Shaziayani membandingkan *Decision Tree*, *Random Forest*, dan *Boosted Regression Tree* pada dataset delapan parameter harian polutan udara di Malaysia. Penelitian menghasilkan, *Random Forest* menunjukkan kinerja paling baik dengan akurasi mencapai 98% (Shaziayani, et al., 2022).

Multilayer Perceptron (MLP) dan *Random Forest* merupakan dua metode dengan pendekatan yang berbeda, persamaan keduanya adalah dapat digunakan

untuk tujuan klasifikasi dan regresi. Pada klasifikasi, kedua algoritma tersebut digunakan ketika target kelas adalah tipe kategoris.

MLP merupakan bagian dari *Artificial Neural Network* yang terdiri dari sejumlah besar elemen pemrosesan yang saling berhubungan yang disebut *neuron*, sedangkan *Random Forest* adalah teknik *ensemble learning* berdasarkan kombinasi dari satu set besar *Decision Tree*.

Berdasarkan kajian yang telah dipaparkan, terdapat peluang penelitian untuk melakukan perbandingan kinerja algoritma *Multilayer Perceptron* dan *Random Forest* dalam kasus klasifikasi kategori Indeks Standar Kualitas Udara di DKI Jakarta melalui nilai akurasi, presisi, *recall*, dan *f1-score* dengan variabel penelitian adalah hasil monitoring konsentrasi polutan partikulat (PM₁₀ dan PM_{2.5}), nitrogen dioksida (NO₂), sulfur dioksida (SO₂), karbon monoksida (CO), dan ozon (O₃) pada tahun 2021. Perbandingan antara metode klasifikasi *Multilayer Perceptron* (MLP) dan *Random Forest* dengan kinerja paling optimal akan digunakan untuk evaluasi prediksi kategori ISPU pada data polutan yang belum diketahui kategorinya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini adalah:

1. Bagaimana membandingkan metode *machine learning Multilayer Perceptron* (MLP) dan *Random Forest* pada kasus klasifikasi kategori Indeks Standar Kualitas Udara (ISPU) di DKI Jakarta?

2. Bagaimana perbandingan nilai *accuracy*, *precision*, *recall*, dan *f1-score* dari metode *Multilayer Perceptron* (MLP) dan *Random Forest* pada kasus klasifikasi kategori Indeks Standar Pencemaran Udara (ISPU) di DKI Jakarta?
3. Bagaimana implementasi antara model klasifikasi dengan metode *Multilayer Perceptron* (MLP) atau *Random Forest* jika dilakukan evaluasi untuk memprediksi nilai kategori ISPU pada data polutan yang belum diketahui kategori ISPU-nya?

1.3 Batasan Masalah

Batasan masalah yang ditentukan dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data Indeks Standar Pencemaran Udara (ISPU) di DKI Jakarta dari Januari sampai dengan Desember 2021.
2. Implementasi metode *Multilayer Perceptron* (MLP) dan *Random Forest* pada kasus klasifikasi kategori Indeks Standar Pencemaran Udara (ISPU) di DKI Jakarta dilakukan secara terpisah, tidak terintegrasi satu sama lain.
3. Jenis polutan yang menjadi variabel penelitian sesuai dengan data hasil pemantauan yaitu partikulat (PM₁₀ dan PM_{2.5}), nitrogen dioksida (NO₂), sulfur dioksida (SO₂), karbon monoksida (CO), dan ozon (O₃).
4. Kategori Indeks Standar Pencemaran Udara (ISPU) sesuai aturan yang berlaku di DKI Jakarta.

5. Penerapan metode *Multilayer Perceptron* dan *Random Forest* dilakukan dengan memanfaatkan *library* yang terdapat pada bahasa pemrograman *python*.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Menerapkan metode *machine learning Multilayer Perceptron* (MLP) dan *Random Forest* pada kasus klasifikasi kategori Indeks Standar Pencemaran Udara (ISPU) di DKI Jakarta.
2. Melakukan analisis perbandingan nilai *accuracy*, *precision*, *recall*, dan *f1-score* pada metode *Multilayer Perceptron* (MLP) dan *Random Forest* untuk mendapatkan model klasifikasi yang paling optimal.
3. Melakukan evaluasi model klasifikasi untuk memprediksi kategori ISPU pada data polutan yang belum diketahui kategorinya, dengan salah satu diantara metode *machine learning Multilayer Perceptron* (MLP) atau *Random Forest*.

1.5 Manfaat Penelitian

Melalui penelitian ini diharapkan dapat memberi manfaat, antara lain:

1. Bagi penulis, penelitian ini merupakan peluang untuk menambah wawasan pengetahuan baik dalam teori maupun praktik dan merupakan salah satu tugas

akademik dalam rangka memenuhi syarat kelulusan di program studi Informatika di Universitas Siliwangi.

2. Bagi pihak yang memiliki peran dalam pengendalian kualitas udara secara langsung, informasi yang didapatkan dari hasil penelitian ini dapat menjadi acuan sebagai salah satu pemanfaatan metode *machine learning* pada kasus klasifikasi kategori Indeks Standar Pencemaran Udara (ISPU) dan dapat diterapkan pada layanan informasi ISPU.
3. Bagi pembaca dan peneliti, penelitian ini dapat menjadi masukan wawasan pengetahuan dan dapat menjadi acuan untuk melakukan penelitian sejenis atau dikembangkan melalui penelitian yang lebih luas.

1.6 Metodologi Penelitian

Metodologi penelitian meliputi beberapa tahapan yang dilakukan dalam proses penelitian yaitu:

1. Studi Pustaka

Studi pustaka yang dilakukan untuk mengumpulkan informasi yang relevan dengan topik yang menjadi objek penelitian. Sumber studi pustaka yang dilakukan berupa buku, *e-book*, jurnal nasional dan beberapa jurnal internasional.

2. Identifikasi Masalah

Identifikasi masalah merupakan proses mendefinisikan masalah serta membuat definisi tersebut menjadi lebih terukur dan dianggap bisa ditemukan jawabannya melalui sebuah penelitian yang dilakukan secara ilmiah.

3. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data harian konsentrasi polutan udara yang diukur dari 5 Stasiun Pemantau Kualitas Udara (SPKU) yang ada di Provinsi DKI Jakarta tahun 2021, data tersedia pada laman Jakarta Open Data yang dapat di akses pada alamat <https://data.jakarta.go.id/dataset/indeks-standar-pencemaran-udara-ispu-tahun-2021>.

4. Praproses Data

Praproses data dilakukan untuk mengubah data mentah yang telah dikumpulkan menjadi informasi yang lebih bersih dan bisa digunakan pada tahap penerapan metode *machine learning*.

5. Proses Klasifikasi

Menerapkan metode *machine learning Multilayer Perceptron* (MLP) dan *Random Forest* pada dataset ISPU DKI Jakarta yang sudah dilakukan praproses.

6. Perbandingan Model Klasifikasi

Perbandingan kinerja model klasifikasi dengan metode *Multilayer Perceptron* (MLP) dan *Random Forest* melalui nilai akurasi, presisi, *recall*, dan *f1-score*.

7. Evaluasi Model Untuk Prediksi Kategori ISPU

Model klasifikasi dengan nilai kinerja model paling optimal akan digunakan untuk evaluasi prediksi kategori ISPU pada data polutan yang belum diketahui kategorinya..

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini diuraikan sebagai berikut:

BAB I PENDAHULUAN

Bab I membahas tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab II memaparkan dasar teori dan melakukan studi pustaka penelitian-penelitian terkait yang berhubungan dengan permasalahan yang diteliti dan menjadi acuan konseptual.

BAB III METODOLOGI

Bab III memaparkan metodologi dan langkah-langkah dalam melakukan penelitian, seperti proses pengumpulan data, praproses data dan tahapan penelitian dalam menerapkan metode metode *Multilayer Perceptron* (MLP) dan *Random Forest*.

BAB IV HASIL DAN PEMBAHASAN

Bab IV memaparkan hasil dan pembahasan dari penelitian, seperti data hasil praproses, hasil implementasi metode *Multilayer Perceptron* (MLP) dan *Random Forest*, hasil perbandingan kinerja model, dan melakukan evaluasi untuk prediksi kategori ISPU.

BAB V KESIMPULAN

Bab V memaparkan kesimpulan yang diperoleh dari hasil penelitian dan saran yang dapat diterapkan dari hasil penelitian untuk menjadi masukan yang berguna bagi penelitian selanjutnya.