

BAB 1

PENDAHULUAN

1.1 Latar Belakang

DKI Jakarta termasuk ke dalam 20 besar kota terpolusi di dunia (IQAir, 2022), yang menunjukkan bahwa pencemaran udara sangat tinggi. Pencemaran udara merupakan salah satu penyebab gangguan kesehatan seperti premature, masalah pada paru-paru, dan lain-lain (Viegi et al., 2020). Pencemaran udara diakibatkan terkontaminasinya udara bersih dengan zat-zat berbahaya yang disebut polutan (Arty, 2005). Unsur-unsur polutan pada negara Asia menggunakan O₃, PM₁₀, PM_{2.5} AQI, NO₂/NO_X, SO₂, CO sebagai indikator penentu polutan udara (Rybarczyk & Zalakeviciute, 2018). Klasifikasi polutan udara dilakukan untuk menentukan kualitas udara yaitu baik, sedang, tidak sehat, sangat tidak sehat, dan berbahaya (Kementrian Lingkungan Hidup dan Kehutanan, 2020).

Pada penelitian (Sethi & Mittal, 2019) dilakukan klasifikasi kualitas udara menggunakan algoritma Decision Trees, dan Naïve Bayes. Hasil menunjukkan tingkat akurasi 95,5 % error 0,45% pada Decision Trees dan akurasi 73,9% error 26% pada Naïve Bayes. Penelitian (Gore & Deshpande, 2017) dilakukan klasifikasi kualitas udara dengan menunjukkan hasil tingkat akurasi algoritma Decision Trees 91,9% dan 86,6% pada Naïve Bayes. Penelitian (Lubis et al., 2020) melakukan peningkatan algoritma k-NN menggunakan *principal component analysis* (PCA) ke dalam tahapan *feature selection* untuk melakukan klasifikasi data kualitas udara, menunjukkan rata-rata akurasi sebesar 87% dengan tingkat error tertinggi 0,11% dan terendah sebesar 0,08%. Penelitian (Singgih, 2020) dalam klasifikasi kualitas udara menunjukkan tingkat akurasi Decision Tree metode Random Forest memiliki tingkat akurasi 81% sedangkan Decision Tree metode CART menunjukkan tingkat akurasi sebesar 75%.

Berdasarkan kajian penelitian, dapat disimpulkan bahwa tingkat akurasi Decision Tree, k-NN, dan Naïve Bayes dalam mengklasifikasi data kualitas udara cukup baik. Bilangan kontinu pada data dapat ditangani oleh Decision Tree namun, tidak cukup baik dalam *overfitting* (Ameer et al., 2019). K-NN memiliki tingkat efektivitas yang baik dalam mengolah data yang cukup besar tetapi memiliki kelemahan yaitu observasi dalam menentukan nilai K terbaik, yang nantinya akan berpengaruh terhadap hasil penambahan data (Aini & Mustafa, 2020).

Pada penelitian ini akan dilakukan pengujian dan perbandingan algoritma Decision Trees, Naïve Bayes, dan k-NN untuk menguji masing-masing algoritma dan juga mendapatkan perbandingan akan model terbaik dalam melakukan klasifikasi kualitas udara menggunakan *dataset* indeks standar pencemaran udara (ISPU) Jakarta Open Data.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah, maka penelitian ini berfokus pada:

1. Bagaimana melakukan klasifikasi *dataset* indeks standar pencemaran udara (ISPU) menggunakan algoritma Decision Tree?
2. Bagaimana melakukan klasifikasi *dataset* indeks standar pencemaran udara (ISPU) menggunakan algoritma K-Nearest Neighbor?
3. Bagaimana melakukan klasifikasi *dataset* indeks standar pencemaran udara (ISPU) menggunakan algoritma Naïve Bayes?
4. Bagaimana perbandingan nilai akurasi dan error tiap algoritma klasifikasi pada *dataset* indeks standar pencemaran udara (ISPU)?

1.3 Tujuan Tugas Akhir

1. Menerapkan algoritma klasifikasi Decision Tree pada *dataset* indeks standar pencemaran udara (ISPU) untuk proses penambangan data.
2. Menerapkan algoritma klasifikasi K-Nearest Neighbor pada *dataset* indeks standar pencemaran udara (ISPU) untuk proses penambangan data.
3. Menerapkan algoritma klasifikasi Naïve Bayes pada *dataset* indeks standar pencemaran udara (ISPU) untuk proses penambangan data.
4. Memaparkan nilai akurasi dan error tiap algoritma klasifikasi pada *dataset* indeks standar pencemaran udara (ISPU), agar pembaca mendapatkan perbandingan tiap algoritma.

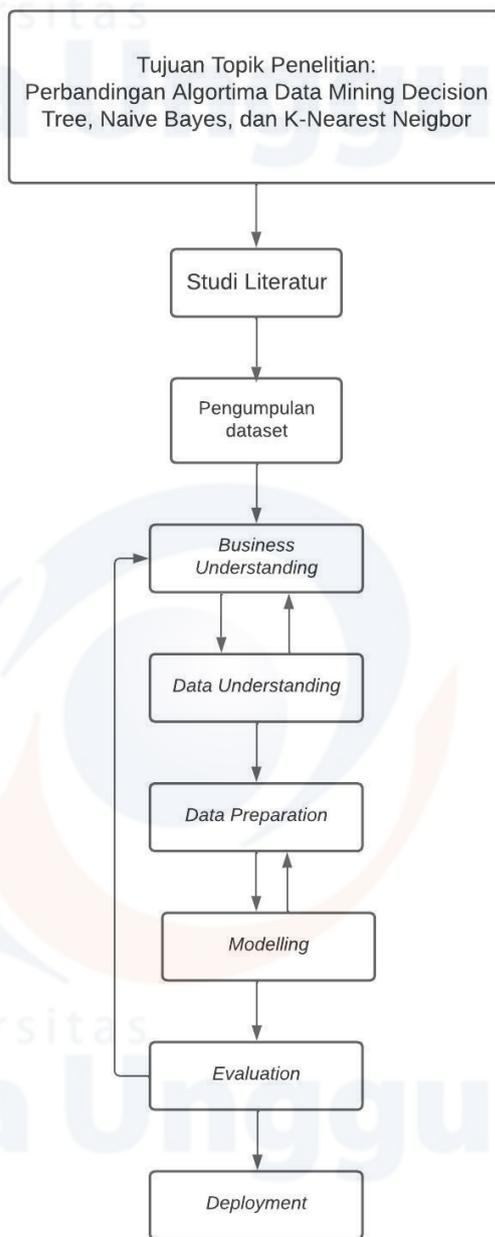
1.4 Manfaat Tugas Akhir

1. Memberikan informasi serta permasalahan yang ada terhadap pembaca akan proses dalam penambangan data.
2. Pembaca mendapatkan gambaran akan tiap metode algoritma dalam proses penambangan data kualitas udara Jakarta Open Data.
3. Mendorong minat pembaca dalam melakukan penambangan data dengan menggunakan *dataset*, algoritma, dan berbagai metode yang berbeda.
4. Peneliti mengimplementasikan ilmu yang di dapat selama perkuliahan, dalam proses penambangan data menggunakan algoritma klasifikasi pada *dataset* indeks standar pencemaran udara (ISPU).

1.5 Lingkup Tugas Akhir

1. Penambangan *dataset* indeks standar pencemaran udara (ISPU) menggunakan algoritma klasifikasi Decision Tree, Naïve Bayes, dan K-Nearest Neighbor dengan *tools* Google Colaboratory.
2. Membandingkan nilai akurasi dan error menggunakan confusion matrix pada *dataset* indeks standar pencemaran udara (ISPU).

1.6 Kerangka Berpikir



Gambar 1-1 Kerangka Berpikir Penelitian

1.7 Sistematika Penulisan Tugas Akhir

Sistematika penulisan laporan tugas akhir ini terbagi menjadi beberapa Bab dan Sub Bab berikut:

BAB I PENDAHULUAN

Pada bab ini berisi penjelasan tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan tugas akhir, kerangka berpikir, dan sistematika penulisan laporan tugas akhir.

BAB II TINJAUAN PUSTAKA

Pada bab ini berisi penjelasan semua materi yang digunakan dalam menyusun laporan tugas akhir.

BAB III METODOLOGI PENELITIAN

Pada bab ini diterangkan tentang Rencana Penelitian, *Standard Process Data Mining (CRISP-DM)*, Perhitungan Algoritma dan Teknik Terkait dalam penelitian Tugas Akhir.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini dilakukan implementasi keseluruhan proses *mining* menggunakan *Google Collaboratory* menggunakan beberapa teknik *preprocessing*, algoritma *Decision Tree*, *Naïve Bayes*, *K-Nearest Neighbor* dan mengevaluasi untuk melihat model terbaik.

BAB V KESIMPULAN DAN SARAN

Pada bab ini dijelaskan kesimpulan dan saran-saran yang berhubungan dengan masalah-masalah yang dibahas untuk Tugas Akhir ini.