

Bab I

Pendahuluan

1.1 Latar Belakang

Berdasarkan jurnal *Estimating the costs of air pollution to the National Health Service and social*, pada tahun 2035 biaya yang dikeluarkan untuk menyembuhkan penyakit yang diakibatkan oleh polusi udara sekitar £18.57 milyar euro. Serta terdapat sekitar 2,5 juta kasus *Non Communicable Disease* (NCD) yang disebabkan oleh polusi udara diprediksi pada tahun 2035 jika PM 2.5 dan NO₂ tetap pada tingkat saat ini yang membuat keadaan menjadi lebih buruk dimasa depan. Pimpin, Retat, Fecht, de Preux, Sassi, Gulliver, Belloni, Ferguson, Corbould, Jaccard and Webber (2018).

Terdapat beberapa zat yang dapat dikategorikan sebagai polusi udara contohnya adalah *particulate matter*(PM) 2.5, *nitrogen monoksida*(NO₂), *Ozone*(O₃), *sulfur dioksida*(SO₂), dsbCastelli, Clemente, Popovič, Silva and Vaneschi (2020). Zat-zat tersebut merupakan penyebab dari terjadinya polusi udara. Mungkin dalam kehidupan sehari-hari zat-zat tersebut dampaknya tidak terlihat tetapi dampaknya baru terlihat ketika jangka panjang. Salah satu polutan terbanyak di Indonesia adalah gas CO yang berasal dari gas buang dari kendaraan baik roda dua maupun roda empatIsmiyati, Marlita and Saidah (2014).

Saat ini sudah terdapat banyak penelitian yang mengusulkan metode pendeteksi polusi udara menggunakan *machine learning*. Namun dari sekian banyak penelitian seperti yang dilakukan oleh Castelli et al. (2020),Bhattacharya and Shahnawaz (2021), kebanyakan dari karya tersebut tidak melakukan perbandingan antara satu metode dengan metode yang lain yang menyebabkan kurangnya pengetahuan tentang algoritma terbaik untuk memprediksi polusi udara. Serta dari ketiga refrensi jurnal di atas masih menggunakan algortima machine learning yang klasik.

Oleh karna itu, pengembangan ensemble learning pada prediksi polusi udara menggunakan masih jarang dilakukan, beberapa penelitian yang sudah melakukan pengembangan ensemble learning seperti pada jurnal Chang, Abimannan, Chiao, Lin and Huang (2020) namun belum banyak jurnal yang sudah menerapkan *ensemble learning* untuk memprediksi polusi udara dikarenakan

kebanayakan jurnal masih menggunakan algoritma *machine learning* klasik.

Dikarenakan hal tersebut maka penulis akan melakukan penelitian tentang study kinerja algoritma *ensemble learning* untuk mendeteksi polusi udara.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah tugas akhir ini adalah sebagai berikut:

1. Bagaimana cara pengembangan *ensemble learning* prediksi polusi udara berdasarkan algoritma terbaik ?
2. Bagaimana cara pengembangan web Prototype pendeteksi polusi udara berbasis *ensemble learning*?
3. Bagaimana mengetahui performansi dari *ensemble learning* yang dikembangkan?

1.3 Pernyataan Masalah

Berdasarkan latar belakang di atas, dapat disimpulkan terdapat permasalahan pada algoritma ekstraksi ciri dan deteksi yang sudah ada sebagai berikut:

1. Pengembangan *ensemble learning* deteksi polusi udara berdasarkan algoritma terbaik masih jarang dilakukan
2. Kurangnya web prototype sebagai validasi dari deteksi polusi udara dengan model *ensemble learning*
3. Kurangnya analisis performansi dari *ensemble learning* untuk memprediksi polusi udara

1.4 Tujuan

Berdasarkan perumusan masalah dan pernyataan masalah, tujuan tugas akhir ini adalah sebagai berikut:

1. Melakukan studi tentang algoritma *ensemble learning* terbaik untuk memprediksi polusi udara
2. Mengembangkan web prototype sebagai validator dari model yang sudah di bangun
3. Melakukan analisis performansi berupa metrik performansi dari *ensemble learning* yang digunakan

1.5 Batasan Masalah

Berikut adalah ruang lingkup yang ada pada penulisan tugas akhir ini :

1. Jenis zat-zat yang diprediksi oleh algoritma machine learning adalah zat-zat yang menyebabkan polusi udara seperti *Ozone*(O₃), *sodium dioksida*(NO₂), *particulate matter 10*(PM₁₀), dan *particulate matter 2.5*(PM_{2.5})
2. Algoritma *machine learning* yang digunakan adalah *k-Nearest Neighbors*(k-NN), *Support Vector Machine*(SVM), dan *XGBoost*
3. Data yang digunakan ialah data yang didapat dari *Air Pollution Prediction and Forecasting* yang berasal dari github Unnikrishnan (2020)

1.6 Hipotesis

1. Algoritma pendeteksi udara yang dibandingkan dalam penelitian ini menghasilkan luaran algoritma terbaik yang dapat digunakan untuk memprediksi polusi udara.
2. Menghasilkan algoritma *ensemble learning* terbaik yang dibangun oleh penulis untuk memprediksi polusi udara.

1.7 Sistematika Penulisan

Tugas Akhir ini disusun dengan sistematika penulisan sebagai berikut :

- **BAB I Pendahuluan.** Bab ini membahas mengenai latar belakang, rumusan masalah, dan tujuan pengerjaan Tugas Akhir ini.
- **Bab II Kajian Pustaka.** Bab ini membahas fakta dan teori yang berkaitan dengan perancangan sistem untuk mendirikan landasan berfikir. Dengan menggunakan fakta dan teori yang dikemukakan pada bab ini penulis menganalisis kebutuhan rancangan arsitektur sistem yang dibangun.
- **BAB III Metodologi dan Desain Sistem.** Bab ini menjelaskan metode penelitian, rancangan sistem dan metode pengujian yang dilakukan dalam penelitian.
- **BAB IV Hasil dan Pembahasan.** Bab ini membahas hasil dari penelitian yang sudah dilakukan menggunakan metode pada bab III
- **BAB V Kesimpulan dan Saran.** Bab ini menjelaskan kesimpulan dari penelitian yang sudah dilakukan serta pemberian saran pada penelitian selanjutnya.