

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pencemaran udara merupakan suatu masalah yang berdampak buruk bagi kehidupan makhluk hidup. Udara yang tercemar akan menimbulkan berbagai macam penyakit. Sehingga perlu dilakukan pengamatan tingkat pencemaran udara pada lingkungan masyarakat (Sastra & Rachman, 2016).

Menteri Kesehatan Republik Indonesia, menjelaskan bahwa Pencemaran udara adalah masuknya atau dimasukkannya zat, energi, dan atau komponen lain ke dalam udara oleh kegiatan manusia, sehingga mutu udara turun sampai ketinggian tertentu yang menyebabkan atau mempengaruhi kesehatan manusia (Menteri Kesehatan Republik Indonesia, 2002).

Kota Jakarta termasuk dalam kota dengan tingkat pencemaran udara yang buruk di dunia. Pemantauan kualitas udara yang dilakukan oleh Green Peace pada tahun 2017 juga menunjukkan bahwa polusi udara pada wilayah Jabodetabek berada jauh di atas ambang batas kesehatan *World Healty Organisasi* (WHO) dan Baku Mutu Udara Ambien Nasional. Di sisi lain, masyarakat Jabodetabek masih memiliki pemahaman yang terbatas mengenai polusi udara dan bahayanya, terutama karena akses terhadap informasi mengenai kualitas udara yang sangat minim. (Green Peace, 2017)

Kualitas data harian Indeks Standart Pencemaran Udara (ISPU) yang dilakukan oleh Dinas Lingkungan Hidup selama tahun 2018 melalui stasiun pemantau pada sejumlah titik di Jakarta mengalami perubahan setiap harinya. Data yang ada belum dapat memberikan pemahaman dan informasi yang memadai kepada masyarakat secara terus-menerus. Pemerintah dituntut harus menyediakan informasi kualitas udara lebih memadai dan selanjutnya menerapkannya beragam kebijakan untuk memenuhi hak warga terhadap udara yang berkualitas (Nurjanah, Siregar, & Kusumaningrum, 2020). Salah satu solusi untuk menyelesaikan adalah dengan teknik komputasi.

Untuk mengetahui indeks pencemaran standar udara, digunakan metode dalam data mining yaitu metode *K-Nearest Neighbor*. Tujuan dari *K-Nearest Neighbor* adalah mengklasifikasikan objek baru berdasarkan atribut data dan *training sample* (Farsiah, Abidin, & Munadi, 2013).

Proses tersebut menggunakan teknik statistik, matematika, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi dan pengetahuan yang bermanfaat dan terkait dengan *database* yang besar.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka perumusan masalah terdapat permasalahan yang dapat dirumuskan yaitu :

1. Bagaimana menerapkan algoritma *K-Nearest Neighbor* pada penentuan indeks standar pencemaran udara di provinsi DKI Jakarta tahun 2021
2. Bagaimana keakuratan hasil dari menggunakan algoritma *K-Nearest Neighbor* untuk penentuan sehat atau tidak sehat standar pencemaran udara di provinsi DKI Jakarta tahun 2021

1.3 Tujuan dan Manfaat

1. Menerapkan algoritma *K-Nearest Neighbor* pada penentuan sehat atau tidak sehat udara di provinsi DKI Jakarta tahun 2021
2. Mengetahui efektifitas penentuan standar pencemaran udara di provinsi DKI Jakarta tahun 2021

1.4 Ruang Lingkup

Agar pengerjaan tugas akhir ini menjadi lebih terarah dan mendapatkan hasil yang lebih spesifik, maka sistem yang dirancang dibatasi pada ruang lingkup pembahasan sebagai berikut:

1. Menggunakan aplikasi WEKA.
2. *Dataset* didapat dari situs *Kaggle.com* dan situs Dinas Lingkungan Hidup Jakarta lingkunganhidup.jakarta.go.id

1.5 State of the art

Tabel 1. 1 *State of the art*

DOI	Jurnal	Kesimpulan	Persamaan	Perbedaan
10.26438 /ijcse/v7i5.12531257 <i>International Journal of Computer Sciences and Engineering Vol.-7, Issue-5, May 2019</i>	<i>Implementati on of Classification Algorithms in Educational Data using Weka Tool</i> (Thangamuthu & Sengottaiyan, 2019)	Penelitian ini menganalisis kumpulan data siswa dan menentukan peringkat kinerja siswa yang membutuhkan perhatian khusus dengan menggunakan alat Weka	➤ Menggunakan <i>Weka Tools</i> saat pemrosesan data	➤ Menggunakan metode algoritma SVM, <i>naïve bayes</i> , LMT, dan SMO
10.1109/ICICT50816.2021.9358642 <i>Proceedings of the Sixth International Conference on Inventive Computation Technologies [ICICT 2021]</i>	<i>Detection of Phishing Websites from URLs by using Classification Techniques on WEKA</i> (Geyik, Erensoy, & Koçyiğit, 2021)	Menjalankan sistem deteksi phishing di WEKA dan menguji efisiensinya menggunakan kumpulan data publik sebagai CatchPhish D3	➤ Menggunakan <i>Weka Tools</i> ➤ Menggunakan metode klasifikasi	➤ Membandingkan 4 algoritma klasifikasi berbeda dan yang memiliki akurasi paling besar adalah <i>Random Forest Algorithm</i>
10.35940 /ijeat.B4202.	<i>Weather Predictions</i>	Penggunaan	➤ Menerapkan	➤ Menggunakan

<p>129219 <i>International Journal of Engineering and Advanced Technology (IJEAT) Volume-9 Issue-2, December, 2019</i></p>	<p><i>using Support Vector Machine and Rough Set (Jayasingh & Mantri, 2019)</i></p>	<p><i>Rough Set Theory</i> untuk memprediksi cuaca pada waktu tertentu untuk tempat tertentu</p>	<p>metode akurasi, dan absolute mean error</p>	<p>metode SVM dan <i>Rough Set Theory</i></p>
<p>10.12962/j24775401.v6i2.4360 <i>International Journal of Computing Science and Applied Mathematics, August 2020</i></p>	<p><i>Comparative Study of KNN, SVM and Decision Tree Algorithm for Student's Performance Prediction (Wiyono, Wibowo, Hidayatullah, & Dairoh, 2020)</i></p>	<p>Membandingkan ketiga Algoritma klasifikasi yaitu K-NN, SVM dan <i>Decision Tree</i> untuk pencegahan kemunduran waktu prediksi kelulusan siswa.</p>	<p>➤ Menggunakan Algoritma K-NN</p>	<p>➤ Membandingkan ketiga Algoritma klasifikasi yaitu K-NN, SVM dan <i>Decision Tree</i> yang dimana SVM mempunyai akurasi paling tinggi pada kasus ini.</p>
<p>10.2139/ssrn.3879610 <i>Proceedings of the International Conference on Innovative Computing & Communication (ICICC), July</i></p>	<p><i>Performance Analysis of Classification Techniques in Data Mining using WEKA (Ratra, Gulia, & Gill, 2021)</i></p>	<p>Menyajikan evaluasi komparatif pengklasifikasi WEKA dalam konteks dataset diabetes</p>	<p>➤ Menggunakan <i>Weka Toolkit</i> ➤ Menggunakan Algoritma K-NN</p>	<p>➤ Membandingkan 7 algoritma klasifikasi berbeda dan yang memiliki akurasi paling besar adalah <i>Logistic</i></p>

2021				<i>Regression</i>
------	--	--	--	-------------------

1.6 Sistematika Penulisan

Tugas Akhir ini disusun berdasarkan sistematika penulisan sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini menjelaskan tentang latar belakang masalah yang akan dibahas dalam penelitian ini, rumusan masalah, tujuan penelitian, ruang lingkup, metodologi penelitian, serta sistematika penulisan tugas akhir.

BAB 2 LANDASAN TEORI

Bab ini berisi semua teori yang dirujuk serta penjelasan singkat tentang metode atau algoritma yang digunakan dalam penelitian.

BAB 3 METODE PENELITIAN

Bab ini berisi langkah-langkah yang akan dilakukan dalam penelitian, data masukan, dan implementasi Algoritma *K-Nearest Neighbor*

BAB 4 PENGUJIAN & ANALISIS HASIL

Bab ini berisi pengujian data menggunakan Python dan WEKA dan hasil dari pengujian berserta analisisnya

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang didapat dari penelitian yang dilakukan serta saran untuk penelitian selanjutnya.