

## **BAB II**

### **TINJAUAN PUSTAKA**

Penelitian mengenai aplikasi *machine learning* di Indonesia, khususnya pada bidang deteksi kualitas udara, sudah pernah dilakukan jauh sebelum penelitian ini disusun. Sebagai contoh, penelitian yang dilakukan oleh Kirono dkk. membahas mengenai pemanfaatan algoritma *Naïve-Bayes* dalam mengklasifikasi tingkat kualitas udara di Jakarta. Data yang dianalisis merupakan data ISPU pada bulan Februari hingga Oktober 2021 dengan format .csv. Hasil penelitian menunjukkan nilai akurasi sebesar 91%, yang dimana rerata dari category sedang memberikan akurasi sebesar 87%, nilai *recall* sebesar 96%, nilai *f1-score* sebesar 91% dan akurasi 96%, nilai *recall* sebesar 86%, *f1-score* 91% untuk kategori tidak sehat. Rata-rata dengan *macro-average* menghasilkan presisi 91%, *recall* 91%, dan *score f1* 91%, sedangkan *weighted-average* menghasilkan akurasi 91%, *recall* 91%, dan *f1-score* sebesar 91% [6]

Penelitian kedua dilakukan oleh Suci dkk. menjabarkan mengenai klasifikasi tingkat pencemaran udara pada sektor industri dengan metode *Random forest*. Jenis data yang digunakan merupakan data kuantitatif. Data yang digunakan bersumber dari data sekunder milik Badan Lingkungan Hidup Kota Medan pada tahun 2018. Data tersebut merupakan hasil pemantauan kualitas emisi sumber tidak bergerak dari berbagai industri yang ada di Kota Medan. Berdasarkan hasil analisa, didapat bahwa tingkat akurasi yang didapat dengan pemodelan *Random Forest* bernilai sebesar 0,95 atau 95% nilai presisi. Selain itu variabel paling berpengaruh dalam klasifikasi tingkat polusi udara pada sektor industri yang berada di kota Medan adalah sulfur dioksida (SO<sub>2</sub>) dan nitrogen dioksida (NO<sub>2</sub>) dan opasitas dan partikulat memiliki variabel terendah [7]

Penelitian ketiga dilakukan oleh Adityo, dkk mengulas mengenai klasifikasi tingkat kualitas udara pada DKI Jakarta. *Dataset* yang digunakan merupakan ISPU tahun 2021 yang diperoleh dari *website* Jakarta Open Data. Data diambil dalam kurun waktu Februari hingga Oktober 2021. Metode *data mining* yang digunakan

untuk penelitian ini adalah klasifikasi, yang berguna dalam memproses informasi data parameter ISPU untuk menyelidiki polusi udara. Model pembelajaran mesin yang digunakan pada klasifikasi yaitu model *Random Forest*. Hasil menunjukkan bahwa klasifikasi menggunakan *Random Forest* mempunyai performa terbaik dengan rerata akurasi tertinggi sebesar 90%. Kategori Sedang memiliki rerata *precision* sebesar 90,9%, *recall* sebesar 89,28%, *f1-score* sebesar 90,9%. Kemudian kategori Tidak Sehat memiliki rerata *precision* 89,09%, *recall* sebesar 90,74%, dan *f1-score* sebesar 89,9% [8].

Penelitian keempat ditulis oleh Adinda dkk, membahas pemanfaatan algoritma *K-Nearest Neighbors* dalam memprediksi kualitas udara. Dataset yang menjadi dasar penelitian berasal dari *website* Open Data Jakarta dengan laman url. Dataset menyangkut informasi seputar kualitas Udara DKI Jakarta mulai bulan Januari hingga Maret 2021. Dataset tersusun atas 450 *record* data yang memiliki 10 atribut dan 1 kelas. Selain itu dataset memiliki ekstensi *.csv*. Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* dapat digunakan untuk memprediksi kualitas udara menurut Air Pollutant Standard Index (ISPU). Peramalan dilakukan dengan menggunakan tujuh fitur yaitu parameter udara antara lain PM10, PM2.5, SO2, CO, O3, NO, dan max, untuk menentukan kualitas udara. Dengan menggunakan matriks konfusi untuk mengevaluasi model algoritme, nilai  $K = 7$  bekerja paling baik dengan akurasi 96%, presisi 92%, *recall* 95%, dan pengukuran *f-score* sebesar 93%. Selain itu, sistem prediksi kualitas udara yang dibangun di atas framework Flask dapat menampilkan hasil prediksi kualitas udara DKI Jakarta [9].

Penelitian kelima dibuat oleh Yuliska dan Syaliman meneliti tentang peningkatan akurasi *K-Nearest Neighbors* pada data Indeks Standar Pencemaran Udara Kota Pekanbaru. Dataset yang digunakan merupakan data ISPU Kota Pekanbaru yang didapatkan dari lab Udara Pemerintah Kota Pekanbaru. Data tersusun atas 5 atribut dengan 4 pengkategorian kelas data (Bagus, Sehat, Tidak Sehat, Sangat Tidak Sehat), yang memiliki *record* sebanyak 992 baris data. Hasil penelitian membuktikan bahwa Metode yang diusulkan ternyata memberikan hasil

akurasi yang lebih baik daripada kNN konvensional, dimana nilai eksak terendah yang ditemukan untuk kNN adalah 93,97% dan tertinggi hanya 95,48% ketika k adalah 3 dan 5, sedangkan metode yang diusulkan dapat menghasilkan nilai akurasi terendah sebesar 95,98 % dan tertinggi 97,99% saat k adalah 3 dan 5. Rata-rata *gain* yang diperoleh pada metode yang diusulkan adalah 2,42% [10].

**Tabel 2. 1 Tabel Pembandingan Tinjauan Pustaka**

Peneliti	Tahun	Algoritma	Data	Metrik Evaluasi	Hasil
[6]	2022	<i>Naïve-Bayes</i>	ISPU Februari-Oktober 2021	Akurasi, <i>recall</i> , presisi, <i>f-score</i>	nilai akurasi sebesar 91%, yang dimana rerata dari category sedang memberikan akurasi sebesar 87%, nilai <i>recall</i> sebesar 96%, nilai <i>f1-score</i> sebesar 91% dan akurasi 96%, nilai <i>recall</i> sebesar 86%, <i>f1-score</i> 91% untuk kategori tidak sehat. Rata-rata dengan <i>macro-average</i> menghasilkan presisi 91%, <i>recall</i> 91%, dan <i>score f1</i> 91%, sedangkan <i>weighted-average</i> menghasilkan akurasi 91%, <i>recall</i> 91%, dan <i>f1-score</i> sebesar 91%

[7]	2021	<i>Random forest</i>	Data sekunder Badan Lingkungan Hidup Kota Medan pada tahun 2018	Akurasi	tingkat akurasi yang didapat dengan pemodelan <i>Random Forest</i> bernilai sebesar 0,95 atau 95% nilai presisi
[8]	2023	<i>Random Forest</i>	Dataset ISPU Jakarta Februari-Oktober 2021	Akurasi, presisi, <i>recall</i> , <i>f-score</i>	<p>klasifikasi menggunakan <i>Random Forest</i> mempunyai performa terbaik dengan rerata akurasi tertinggi sebesar 90%.</p> <p>Kategori Sedang memiliki rerata <i>precision</i> sebesar 90,9%, <i>recall</i> sebesar 89,28%, <i>f1-score</i> sebesar 90,9%.</p> <p>Kemudian kategori Tidak Sehat memiliki rerata <i>precision</i> 89,09%, <i>recall</i> sebesar 90,74%, dan <i>f1-score</i> sebesar 89,9%</p>
[9]	2022	<i>K-Nearest Neighbors</i>	Dataset Kualitas Udara DKI Jakarta Januari-Maret 2021	Akurasi, presisi, <i>recall</i> , <i>f-score</i>	<p>nilai <math>K = 7</math> bekerja paling baik dengan akurasi 96%, presisi 92%, <i>recall</i> 95%, dan pengukuran <i>f-score</i> sebesar 93%.</p>

[10]	2020	<i>K-Nearest Neighbors</i>	Data ISPU Kota Pekanbaru	Akurasi, rerata gain	<p>Metode yang diusulkan ternyata memberikan hasil akurasi yang lebih baik daripada kNN konvensional, dimana nilai eksak terendah yang ditemukan untuk kNN adalah 93,97% dan tertinggi hanya 95,48% ketika k adalah 3 dan 5, sedangkan metode yang diusulkan dapat menghasilkan nilai akurasi terendah sebesar 95,98 % dan tertinggi 97,99% saat k adalah 3 dan 5. Rata-rata gain yang diperoleh pada metode yang diusulkan adalah 2,42%.</p>
------	------	----------------------------	--------------------------	----------------------	---