

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Pencemaran Udara**

Pencemaran udara bersumber dari aktifitas alam dan juga kebanyakan dari aktifitas manusia. Pencemaran udara tersebut dapat mempengaruhi keseimbangan kualitas udara normal dan mengakibatkan gangguan terhadap kelangsungan hidup makhluk hidup yang ada di muka bumi (Ismiyati et al., 2014).

Pencemaran udara semakin memburuk seiring dengan kemajuan teknologi, dimana dengan kemajuan teknologi sehingga sumber penghasil polusi udara semakin meningkat. Polusi udara memberikan kontribusi, untuk sebagian besar kontaminasi pada makanan dan air, yang dikonsumsi dalam beberapa kasus rute utama asupan polutan (Msy, 2016). Baku mutu udara ambien merupakan suatu ukuran pada batas atau kadar zat, energi, dan unsur pencemar yang ditenggang keberadaannya dalam udara ambien (Kurniawati, Nurullita, 2017).

Sumber bahan pencemar udara menentukan jenis bahan pencemarnya. Adapun jenis parameter pencemaran udara, yang mana meliputi (Sugiarti, 2015):

- a Nitrogen Dioksida (NO<sub>2</sub>), adalah gas yang bersifat racun yang dapat menimbulkan gangguan pada sistem pernapasan manusia.
- b Sulfur Dioksida (SO<sub>2</sub>) yang ada di udara dapat menyebabkan iritasi saluran pernapasan dan kenaikan sekresi mukosa.

- c Karbon Monoksida (CO) menyebabkan keracunan dapat ditandai dari keadaan ringan berupa pusing, sakit kepala, dan mual.
- d Partikulat Matter (PM10), yaitu partikel udara dalam wujud padat yang berdiameter kurang dari 10  $\mu\text{m}$  yang biasanya disebut dengan PM10 atau dikenal dengan partikel debu.
- e Ozon (O<sub>3</sub>), yaitu salah satu pencemar udara yang terus meningkat konsentrasinya. Ozon pada konsentrasi 0,3 ppm dapat menyebabkan iritasi terhadap hidung dan tenggorokan.

## 2.2 Kualitas Udara ISPU

Kualitas Udara Indeks standar pencemar udara (ISPU) adalah laporan kualitas udara kepada masyarakat untuk memberikan gambaran kepada masyarakat tentang kondisi udara atau bagaimana prediksi pencemaran yang akan terjadi. ISPU juga digunakan sebagai bahan pertimbangan pemerintah pusat dan pemerintah daerah dalam melaksanakan pengelolaan dan pengendalian pencemaran udara yang terjadi (Aini et al., 2019). ISPU terbagi menjadi lima kategori, yaitu baik, sedang, tidak sehat, sangat tidak sehat, dan berbahaya (Aljuaid & Alwabel, 2019).

Perhitungan nilai serta penomoran ISPU itu sendiri menggunakan metode perhitungan yang telah ditetapkan melalui Keputusan Kepala Badan Pengendalian Dampak Lingkungan No. KEP-107/KABAPEDAL/11/1997 (Sianipar, 2017), yaitu:

$$I = \frac{Ia - Ib}{Xa - Xb} (Xx - Xb) + Ib$$

Dimana:

$I = \text{ISPU}$

$I_a = \text{ISPU batas atas}$

$I_b = \text{ISPU batas bawah}$

$X_a = \text{Ambien batas atas}$

$X_b = \text{Ambien batas bawah}$

$X_x = \text{Ambien hasil pengukuran}$

Untuk nilai batas atas dan batas bawah nilai ISPU dapat dilihat pada Tabel 2.1. nilai ISPU dihitung untuk semua parameter yang terukur, sehingga diperoleh nilai ISPU untuk masing-masing parameter kualitas udara. Nilai yang diambil sebagai nilai akhir ISPU untuk pengukuran kualitas udara yang terjadi adalah salah satu nilai ISPU yang tertinggi dari hasil perhitungan terhadap semua parameter kualitas udara. Semakin tinggi nilai ISPU mengindikasikan kualitas udara yang buruk dan berdampak buruk bagi kesehatan (Kurniawan, 2018). Dampak buruk dari hal tersebut dapat mengenai populasi orang sehat (Hermawan et al., 2016).

Tabel 2.1 Kualitas Udara ISPU

Nilai ISPU	Level Kesehatan	Keterangan
0 – 50	Baik	Tingkat kualitas udara yang tidak memberikan efek bagi kesehatan manusia atau hewan dan tidak

		berpengaruh pada tumbuhan, bangunan atau nilai estetika.
51 – 100	Sedang	Tingkat kualitas udara yang tidak berpengaruh pada kesehatan manusia ataupun hewan tetapi berpengaruh pada tumbuhan yang sensitif dan nilai estetika.
101 - 199	Tidak Sehat	Tingkat kualitas udara yang bersifat merugikan pada manusia ataupun kelompok hewan yang sensitif atau bisa menimbulkan kerusakan pada tumbuhan ataupun nilai estetika.
200 – 299	Sangat Tidak Sehat	Tingkat kualitas udara yang dapat merugikan kesehatan pada sejumlah segmen populasi yang terpapar.
300 - 500	Berbahaya	Tingkat kualitas udara berbahaya yang secara umum dapat merugikan kesehatan yang serius pada populasi.

### 2.3 Machine Learning

*Machine Learning* adalah bagian dari *artificial intelligence* yang mengadopsi prinsip dari ilmu komputer dan statistik untuk membuat model yang merefleksikan pola-pola data. Model tersebut dilatih dengan menggunakan berbagai algoritma dalam pendekatan *machine learning* sehingga bisa melakukan klasifikasi terhadap entitas bencana (Handayani et al., 2020).

#### 1. Regression

Data yang ada diberikan *real value*, *numerical* atau *floating point*, agar dapat mencoba mendeteksi harga saham di kemudian hari.

#### 2. Classification (Discrete/ Category)

Data yang ada diberikan label atau kategori, agar dapat diambil keputusan berdasarkan label/ kategori tersebut.

Adapun tahapan-tahapan yang harus dilakukan pada proses *machine learning* yaitu, sebagai berikut:

1. Hal pertama yang harus disiapkan adalah data. Pada *machine learning*, data biasanya terdiri dari data *training* dan data *testing*.
2. Data *training* adalah data untuk melatih algoritma untuk menemukan model yang cocok.
3. Data *testing* adalah data untuk menguji dan menentukan kinerja model yang diperoleh pada tahap pengujian.

## **2.4 Klasifikasi**

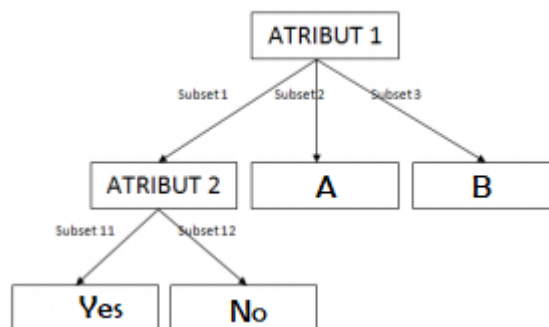
Klasifikasi merupakan suatu proses menemukan kumpulan pola atau fungsi yang mendeskripsikan serta memisahkan kelas data yang satu dengan yang lainnya untuk menyatakan objek tersebut masuk pada kategori tertentu yang sudah dikategorikan. Klasifikasi adalah bentuk analisis data yang mengekstrak model yang menggambarkan kelas data (Novandya & Oktaria, 2017).

Metode klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan

data ke dalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Model itu sendiri bisa berupa aturan “jika-maka”, berupa pohon keputusan, atau formula matematis (Iriadi & Nuraeni, 2016).

### 2.5 Decision Tree

*Decision tree* adalah *flowchart* seperti pohon dimana setiap *node* menunjukkan suatu test pada suatu atribut, tiap branch merepresentasikan hasil dari test tersebut, dan *leaf node* menunjukkan kelas-kelas atau distribusi kelas. *Decision tree* berguna untuk mengeksplorasi data yang sudah melewati tahap preprocessing dan menemukan model yang tersembunyi dari data dengan sebuah target variabel sehingga dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan *record* yang lebih kecil dengan memperhatikan variabel tujuannya (Zega, 2014). Dapat dilihat pada Gambar 2.3.



Gambar 2.1 Diagram Alir *Decision Tree*

*Decision tree* merupakan salah satu metode untuk mengklasifikasikan data (Romadloni et al., 2019). Menggunakan konsep diagram alir yang mirip dengan struktur

pohon, dimana internal *node* menotasikan atribut yang diuji, setiap cabang nya merepresentasikan hasil dari atribut tes tersebut, dan *leaf node* merepresentasikan kelas- kelas tertentu atau distribusi kelas-kelas. Hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari *record* yang baru (Elisa, 2017). Pada umumnya ada beberapa ciri kasus yang cocok untuk diterapkan dengan algoritma ini, antara lain:

- a. Data dinyatakan dengan pasangan atribut dan nilainya misalnya atribut *temperature* dan nilainya adalah dingin.
- b. Label/ *output* data biasanya bernilai diskrit. Outputnya bias bernilai ya atau tidak, sakit atau tidak sakit, diterima atau ditolak.
- c. Data mempunyai *missing value*. Misalkan untuk beberapa data, nilai dari suatu atribut nya tidak diketahui. Maka dalam keadaan ini *decision tree* masih mampu member solusi yang baik.

## **2.6 K-Nearest Neighbor**

*K-Nearest Neighbor* (KNN) adalah kelompok pembelajaran berbasis instans. Algoritma *K-Nearest Neighbor* merupakan metode yang berfungsi untuk mengklasifikasikan objek berdasarkan data pembelajaran yang merupakan jarak terdekat dengan objek yang diuji (Jadhav & Channe, 2016).

Untuk menentukan jarak antara dua titik yaitu titik pada data latih (x) dan titik pada data uji (y) digunakan rumus *Euclidean*, seperti terlihat pada persamaan berikut:

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Dengan  $d$  adalah jarak antar titik pada data latih  $x$  dan data titik uji  $y$  yang akan diklasifikasikan, dimana  $x = x_1, x_2, x_3, \dots, x_i$  dan  $y = y_1, y_2, y_3, \dots, y_i$  dan  $I$  merepresentasikan nilai atribut dan  $n$  adalah dimensi atribut. Pada fase pelatihan, algoritme ini hanya menyimpan vektor fitur dan mengklasifikasikan sampel data pelatihan. Pada fase ini, fitur yang sama dihitung untuk menguji data yang diklasifikasikan sebagai belum diketahui Jarak dari vektor baru ke mana semua vektor sampel pelatihan dihitung dan jumlah  $k$  buah terdekat diambil.

### ***2.7 K-Fold Cross Validation***

*K-Fold Cross Validation* digunakan untuk menguji pola klasifikasi. Dalam *k-fold cross validation* data dibagi menjadi  $k$  bagian,  $D_1, D_2, \dots, D_k$ , dan masing-masing  $D$  memiliki jumlah data yang sama. Menghitung nilai akurasi dapat dilakukan dengan menggunakan persamaan (Wahyuningsih & Utari, 2018):

$$\text{Akurasi} = \frac{\text{Jumlah Klasifikasi Benar}}{\text{Jumlah Data Uji}} \times 100\%$$



## 2.8 Penelitian Terkait

Tabel 2.2 *State of the Art*

No	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
1	Fiqhri Mulianda Putra, Imas Sukaesih Sitanggang	2020	<i>Classification Model Of Air Quality In Jakarta Using Decision Tree Algorithm Based On Air Pollutant Standard Index</i>	Kualitas udara harian perlu diukur dan diklasifikasikan secara akurat. Hasil klasifikasi yang akurat akan membantu pemerintah dalam pengambilan kebijakan. Tujuannya untuk mengendalikan pencemaran untuk mendapatkan standar kualitas udara yang dapat bermanfaat bagi kelangsungan hidup khususnya di Jakarta. Parameter yang mempengaruhi kualitas udara di Jakarta menggunakan algoritma C5.0 dan <i>Random Forest</i> berdasarkan kategori Indeks Standar Pencemaran Udara (ISPU). Algoritma klasifikasi yang digunakan adalah C5.0 dan <i>Random Forest</i> yang dikategorikan	Hasil penelitian menunjukkan bahwa model <i>Random Forest</i> memberikan akurasi terbaik dalam prediksi pada dataset 2017 dengan hasil akurasi 99,74%, 99,22%, dan 99,97% dengan 1.412 data latih dan 389 data uji. Dibandingkan dengan model <i>rule-based</i> C5.0 dan model <i>decision tree</i> dari C5.0. Model <i>rule-based</i> dari C5.0 memberikan akurasi prediksi terbaik pada dataset 2018 dengan hasil akurasi 98,28%, 98,85%, dan 97,42% dengan 1.439 data latih dan 349 data uji. Dibandingkan dengan model <i>decision tree</i> C5.0 dan model <i>Random Forest</i> . Variabel

				dalam model <i>Decision Tree</i> . C5.0 juga menghasilkan model berbasis aturan.	terpenting untuk mengklasifikasi kualitas udara adalah <i>Ozone</i> (O) dan kualitas udara di Jakarta didominasi oleh kategori <i>Moderate</i> pada tahun 2017 dan 2018.
2	Krishna Chaitanya Atmakuri, Dr. K V Prasad	2020	<i>A Comparative Study On Prediction Of Indian Air Quality Index Using Machine Learning Algorithms</i>	Menganalisis sampel data berdasarkan algoritma yang berbeda untuk memprediksi akurasi tingkat AQI berdasarkan pembacaan NO <sub>2</sub> , SO <sub>2</sub> dan SPM. Sistem ini mencoba untuk memprediksi keakuratan level AQI. Algoritma <i>Optimized Bayesian Network</i> dibandingkan dengan algoritma <i>Decision Tree</i> , <i>Random Forest</i> yang berbeda untuk memprediksi level AQI. Analisis kualitas udara berdasarkan kumpulan data yang terdiri dari kondisi atmosfer harian di negara tersebut dan berikan gambaran tentang algoritme mana yang paling cocok untuk memprediksi kualitas udara di masa mendatang.	Hasil penelitian membuktikan bahwa algoritma <i>Optimized Bayesian Network</i> berjalan dengan akurasi terbaik dengan hasil akurasi 99.63. Sementara <i>Naive Bayes</i> , <i>Decision Tree</i> dan <i>Random Forest</i> dengan hasil akurasi 81.95, 99.39 dan 99. Algoritme ini membantu memprediksi tingkat AQI secara akurat, sehingga memungkinkan otoritas pemerintah dan Badan Pengawas Pencemaran untuk mengambil keputusan bijak tentang pengendalian pencemaran di wilayah yang terkena dampak.

3	Siti Nurjanah, Amril Mutoi Siregar, Dwi Sulistya Kusumaningrum	2020	Penerapan Algoritma <i>K – Nearest Neighbor</i> (KNN) Untuk Klasifikasi Pencemaran Udara Di Kota Jakarta	Sumber pencemaran udara dapat disebabkan dari semakin banyaknya kendaraan bermotor dan pembangunan gedung yang tidak memperhatikan keseimbangan alam. Di Kota Jakarta Indeks Standar Pencemaran Udara (ISPU) yang dilakukan oleh Dinas Lingkungan Hidup melalui stasiun pemantau setiap hari selalu mengalami perubahan kategori pencemaran udara. Data yang ada belum mampu memberikan informasi yang memadai kepada masyarakat Jakarta. Sehingga kesadaran masyarakat terhadap bahaya polusi udara belum maksimal. Salah satu cara untuk mengolah data ISPU harian di kota Jakarta adalah dengan teknik klasifikasi data mining menggunakan algoritma <i>K-Nearest Neighbor</i> . Sehingga diharapkan dapat membantu memberikan informasi	Hasil klasifikasi menggunakan algoritma KNN untuk menentukan data ISPU pada kota Jakarta. Maka, dapat ditarik kesimpulan bahwa data latih sebanyak 304 data dan satu data uji dapat menghasilkan kategori klasifikasi baru untuk memprediksi kualitas udara pada bulan berikutnya. Selanjutnya hasil kategori mayoritas $k=7$ yang ada menunjukkan bahwa kualitas udara di bulan berikutnya adalah “SEDANG” dan hasil penelitian ini menghasilkan akurasi sebesar 95.78%. Hasil klasifikasi pencemaran udara di Kota Jakarta, diharapkan dapat membantu Dinas Lingkungan Hidup Kota Jakarta untuk memberikan informasi yang lebih memadai kepada masyarakat. Sehingga masyarakat dapat melakukan tindakan pencegahan
---	--	------	--	---	--

				yang bermanfaat bagi masyarakat dan bagi Dinas Lingkungan Hidup.	terhadap dampak buruk pencemaran udara ketika beraktifitas diluar ruangan.
4	Dewi Sartika, Dana Indra Sensue	2017	Perbandingan Algoritma Klasifikasi <i>Naïve Bayes</i> , <i>Nearest Neighbor</i> dan <i>Decision Tree</i> Pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian	Salah satu teknik data mining yang sering digunakan yaitu teknik klasifikasi. Klasifikasi termasuk tipe <i>supervised learning</i> . Terdapat 5 kategori klasifikasi yaitu berbasis statistik, berbasis jarak, berbasis pohon keputusan, berbasis jaringan syaraf dan berbasis aturan. Tiap kategori memiliki banyak algoritma, contoh algoritma klasifikasi yang sering digunakan yaitu <i>naïve bayes</i> , <i>nearest neighbour</i> dan <i>decision tree</i> . Dari ketiga algoritma tersebut dilakukan proses klasifikasi pada studi kasus pemilihan pola pakaian.	Hasil perbandingan algoritma klasifikasi <i>decision tree</i> memiliki tingkat akurasi paling tinggi dibandingkan <i>naïve bayes</i> dan <i>nearest neighbour</i> yaitu mencapai 75,6%. Mode pengujian menggunakan <i>perctage split</i> dan algoritma <i>decision tree</i> yang digunakan adalah algoritma J48 dengan <i>pruned</i> yang model pohon keputusannya memiliki 166 daun dan 225 ukuran pohon.
5	Yogiek Indra Kurniawan	2018	Perbandingan Algoritma <i>Naïve Bayes</i> dan C,45	Dilakukan Perbandingan dua Algoritma dalam hal <i>Precision</i> , <i>recall</i> dan <i>Accuracy</i> dalam 4 studi kasus, yaitu kasus penerimaan “Kartu	Hasil kasus penentuan penerimaan Kartu Indonesia Sehat, kedua algoritma sama-sama efektif untuk digunakan. Hasil asus penentuan

			<p>Dalam Klasifikasi Data Mining</p>	<p>Indonesia Sehat”, penentuan pengajuan kartu kredit di sebuah bank, penentuan usia kelahiran dan penentuan kelayakan calon anggota kredit pada koperasi. Diimplementasikan dalam sebuah aplikasi yang menerapkan algoritma <i>Naïve Bayes</i> dan C.45. Aplikasi tersebut telah diuji dengan <i>blackbox</i> dengan hasil yang valid dan dapat mengimplementasikan kedua algoritma dengan benar.</p>	<p>pengajuan kredit disetujui bank, algoritma C.45 memberikan hasil yang lebih baik dari pada algoritma <i>Naïve Bayes</i>. Hasil kasus penentuan usia kelahiran, algoritma <i>Naïve Bayes</i> memberikan klasifikasi yang lebih baik dari pada algoritma C.45. Sedangkan pada kasus penentuan kelayakan calon anggota kredit pada koperasi, algoritma <i>Naïve Bayes</i> memberikan nilai yang lebih baik pada <i>precision</i>, tapi untuk <i>recall</i> dan <i>accuracy</i>, algoritma C.45 memberikan hasil yang lebih baik.</p>
6	Devi Yunita	2017	<p>Perbandingan Algoritma <i>K-Nearest Neighbor</i> Dan <i>Decision Tree</i> Untuk</p>	<p>Perbandingan Algoritma <i>K-Nearest Neighbor</i> dan <i>Decision Tree</i> untuk Risiko Kredit Kepemilikan Mobil Dalam penelitian analisa kredit ini menggunakan perbandingan Algoritma <i>K-nearest neighbor</i> (K-NN) yang</p>	<p>Analisa kredit menggunakan Perbandingan Algoritma <i>K-Nearest Neighbor</i> dan Metode <i>Decision Tree</i> dapat meningkatkan tingkat ketelitian bagi Kredit Pemilikan Mobil (KPM) dalam menyeleksi</p>

			Penentuan Risiko Kredit Kepemilikan Mobil	merupakan penelitian menggunakan metode dengan mencari kedekatan antara kriteria kasus baru dengan kriteria kasus lama berdasarkan kriteria kasus yang paling mendekati, dan menggunakan Metode <i>Decision tree</i> yang merupakan metode yang ada pada teknik klasifikasi dalam data mining. Hasil penelitian dengan menggunakan aplikasi Rapid Miner menunjukkan bahwa Algoritma <i>K-Nearest Neighbor</i> (K-NN) memiliki nilai akurasi yang lebih baik	konsumen baru yang akan melakukan kredit mobil. b. Pada hasil penelitian menunjukkan penggunaan Algoritma <i>K-Nearest Neighbor</i> lebih akurat dalam penentuan kelayakan konsumen dengan nilai akurasi 98,18%. Nilai keakuratan ini bertujuan untuk menghindari terjadinya kredit macet.
7	Nurul Aini, Rima Ruktiari, M Riyaldi Pratama, A. Fitrah Buana	2019	Sistem Prediksi Tingkat Pencemaran Polusi Udara dengan Algoritma <i>Naïve Bayes</i> di Kota Makassar	Salah satu yang mempengaruhi masifnya pertumbuhan kendaraan di kota Makassar. Untuk itu maka diperlukan satu sistem yang dapat memprediksi tingkat polusi udara di masa yang akan datang.. Klasifikasi dalam data mining bekerja pada data historis atau data sejarah. Data historis disebut data latihan atau <i>training</i> data.	Hasil pengujian akurasi, bahwa ada beberapa daerah yang memiliki tingkat pencemaran yang tidak sama dengan lokasi pengambilan data sehingga untuk mengambil keputusan status polusi untuk Kota Makassar tidak dapat merata. Hasil pengujian Teknologi <i>Acceptance Model</i> diperoleh

				<p><i>Naïve bayes</i> metode klasifikasi yang memiliki tingkat akurasi lebih baik dari metode klasifikasi lainnya. Sistem ini mampu mengklasifikasi data polusi dikota makassar dan berdasarkan pengujian <i>whitebox</i> dan pengujian teknologi <i>acceptance</i> model dimana semua variabel memenuhi nilai syarat <i>cornbach alpha</i> diatas 0.60 dan ada beberapa faktor yang mempengaruhi tingkat polusi beberapa bagian di kota Makassar.</p>	<p><i>instrument</i> variabel TAM untuk X1 nilai <i>cornbach alpha</i> sebesar 0.79, X2 sebesar 0.78 dan Y sebesar 0.81, dimana semua variabel memenuhi syarat yaitu lebih besar dari 0.60. Sehingga dapat dikatakan <i>instrument</i> penilaian yang digunakan <i>reliable</i>.</p>
8	Kostandina Veljanovska, Angel Dimoski	2017	<p><i>Machine Learning Algorithms in Air Quality Index Prediction</i></p>	<p>Polusi udara di Republik Makedonia dianggap sebagai masalah serius karena nilai terukur dari parameter kualitas udara berkali-kali lipat di atas nilai batas yang ditentukan untuk melindungi kesehatan manusia. Pada tahap pertama dalam memprediksi indeks kualitas udara untuk membantu memperbaiki keadaan.</p>	<p>Algoritma yang paling akurat untuk analisis dan klasifikasi hasilnya adalah NN dengan akurasi maksimum 92,9%, sedangkan algoritma KNN memiliki akurasi maksimum 90,5% dan DT algoritma memiliki akurasi maksimum 81.0%. Algoritma NN berisi 3 lapisan: masukan, 1 tersembunyi dan lapisan keluaran.</p>

				<p>Mengembangkan tiga pengklasifikasi berbeda berdasarkan algoritma yang berbeda dan menggunakan kumpulan data yang didasarkan pada model situs web resmi Kementerian Lingkungan Hidup dan perencanaan fisik Makedonia. Dilakukan perbandingan antara tiga algoritma <i>machine learning</i>, <i>neural network</i>, <i>k-nearest neighbor</i> dan <i>decision tree</i>.</p>	<p>Lapisan masukan berisi 6 atribut masukan, lapisan tersembunyi berisi 10 neuron dan lapisan keluaran berisi 3 kelas.</p> <p>Pada algoritma k-NN dilakukan beberapa kombinasi untuk mendapatkan akurasi tertinggi dengan nilai tetangga terdekat yang berbeda (k). Nilai berada dalam interval saat k = 1 hingga k = 21. Kesimpulan bahwa algoritma akurasi terbesar bertahan ketika k = 8 dan memiliki metrik kosinus.</p>
9	Madhuri VM, Samyama Gunjal GH, Savitha Kamalapurkar	2020	<i>Air Pollution Prediction Using Machine Learning Supervised Learning Approach</i>	<p>Konsentrasi polutan udara di udara ambien diatur oleh parameter meteorologi seperti kecepatan angin atmosfer, arah angin, kelembaban relatif, dan suhu. Teknik sebelumnya seperti Probabilitas, Statistik, dll. Dengan kebutuhan untuk memprediksi kelembaban relatif udara dengan mempertimbangkan berbagai</p>	<p>Hasil prediksi <i>machine learning</i> yang diawasi menggunakan algoritma LR, SVM, DT dan RF adalah 6.01, 3.89, 1.34 dan 0.84. Hasil penelitian menunjukkan bahwa prediksi AQI yang diperoleh melalui RF cukup menjanjikan yang dianalisis dengan hasil.</p>



				parameter seperti CO, Tin oksida, hidrokarbon non logam, Benzene, Titanium, NO, Tungsten, Indium oksida, Temperatur dll, pendekatan menggunakan <i>Linear Regression</i> (LR), <i>Support Vector Machine</i> (SVM), <i>Decision Tree</i> (DT), <i>Random Forest Method</i> (RF) untuk memprediksi kelembaban relatif udara dan menggunakan <i>Root Mean Square Error</i> untuk memprediksi keakuratannya.	
10	Adhika Novandya, Isni Oktria	2017	Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi	Laporan cuaca secara umum sering disebut prakiraan cuaca adalah pemanfaatan ilmu pengetahuan dan teknologi untuk memperkirakan atmosfer bumi pada saat akan datang ke suatu tempat tertentu. Data yang digunakan untuk penelitian yang diperoleh dari cuaca dunia online, adalah situs yang menyediakan data dan informasi kondisi cuaca setiap	Akurasi dari algoritma klasifikasi C4.5 menghasilkan nilai sebesar 88.89% yang telah dibuktikan melalui program yang dibuat. Pengembangan pekerjaan yang akan datang dapat mempertimbangkan tidak hanya nilai <i>accuracy</i> dan kappa dari algoritma tersebut, tetapi

				<p>hari. Data yang digunakan memiliki interval waktu setiap tiga jam mulai dari 12 Agustus 2016 pukul 01.00 sampai dengan 20 Agustus 2016 pukul 22.00. Penelitian bertujuan untuk mendapatkan klasifikasi pola cuaca dengan klasifikasi algoritma data mining yaitu algoritma c4.5.</p>	<p>memperhatikan nilai AUC yang dihasilkan..</p>
11	M Mohanapriy, Mrs J Lekha	2018	<p><i>Comparative study between decision tree and knn of data mining classification technique</i></p>	<p>Data Mining adalah metode yang sangat bermanfaat untuk menganalisis data penting. Banyak Peneliti dan organisasi menggunakan data mining untuk mengekstrak pengetahuan yang berguna mengenai kebutuhan mereka. Penambangan data memiliki banyak teknik. Misalnya, Klasifikasi, <i>Clustering</i>, Regresi, <i>Association</i>, <i>Summarization</i>, <i>Time-series</i> dll. Setiap teknik memiliki beberapa algoritma seperti klasifikasi memiliki pohon keputusan, <i>Naïve Bayes</i>, <i>Neural Networks</i> dan sebagainya dan</p>	<p>Klasifikasi <i>Decision tree</i> dan algoritma KNN. Dari topik terakhir, berikut dibahas perbandingan antara DT dan KNN. Ini menunjukkan bahwa algoritma DT merupakan algoritma yang lebih mudah jika dibandingkan dengan KNN dan juga lebih akurat. DT digunakan untuk mempartisi data untuk mendapatkan hasil yang akurat, tetapi di KNN digunakan untuk menemukan nilai yang sama dari data tersebut. Masing-masing dan setiap algoritme memiliki</p>

				<p><i>Clustering</i> memiliki <i>K-means</i> dan seterusnya. Studi perbandingan antara Algoritma Pohon Keputusan dan Algoritma <i>K-Nearest Neighbor Algorithm of Classification</i>.</p>	<p>kelebihan dan kekurangannya sendiri dan tidak pernah semua algoritme memenuhi semua kriteria dan persyaratan.</p>
12	<p>Firman tempola, Miftah Muhammad, Amal Khairan</p>	2018	<p>Perbandingan Klasifikasi Antara KNN dan <i>Naïve Bayes</i> Pada Penentuan Status Gunung Berapi Dengan <i>K-Fold Cross Validation</i></p>	<p>Menentukan status gunung berapi, pusat vulkanologi dan mitigasi bencana geologi yang dilakukan dengan pengamatan visual dan faktor kegempaan. Dalam klasifikasi aktivitas gunung berapi ada 5 kriteria yang digunakan, yaitu empat faktor kegempaan (gempa vulkanik dangkal, gempa tektonik jauh, gempa vulkanik dalam, gempa hembusan) dan satu status sebelumnya. Diklasifikasikan menggunakan metode K-NN dan <i>Naïve Bayes</i>. Kemudian divalidasi menggunakan <i>K-Fold Cross Validation</i>.</p>	<p>Hasil penelitian yang dibagi kedalam 3 fold diperoleh perbandingan akurasi system dengan rata-rata metode K-NN 63,68% dengan standar deviasi 7,47%. Sedangkan rata-rata metode <i>Naïve Bayes</i> 79,71% dengan standar deviasi 3,55%. Dari perbandingan klasifikasi menggunakan metode KNN dan <i>Naïve Bayes</i> disimpulkan penggunaan jarak <i>Naïve Bayes</i> lebih dekat dibandingkan K-NN.</p>

**2.8 Matriks Penelitian**

Tabel 2.3 Tabel Matriks Penelitian

No.	Penulis/Tahun	Judul	Ruang Lingkup								
			Algoritma/Metode			Tujuan		Objek			
			Naïve Bayes	KNN	Decision Tree	Klasifikasi	Prediksi	ISPU	AQI	Pola Pakai an	Lain-nya
1.	(Putra & Sitanggang, 2020)	<i>Classification Model Of Air Quality In Jakarta Using Decision Tree Algorithm Based On Air Pollutant Standard Index</i>	-	-	√	-	-	√	-	-	-
2.	(Atmakuri & Prasad, 2020)	<i>A Comparative Study On Prediction Of Indian Air Quality Index Using Machine Learning Algorithms</i>	√	-	√	-	√	-	√	-	-
3.	(Nurjanah et al., 2020)	<i>Penerapan Algoritma K – Nearest Neighbor (KNN) Untuk Klasifikasi Pencemaran Udara Di Kota Jakarta</i>	-	√	-	√	-	√	-	-	-

4.	(Sartika & Indra, 2017)	Perbandingan Algoritma Klasifikasi <i>Naïve Bayes</i> , <i>Nearest Neighbor</i> dan <i>Decision Tree</i> Pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian	√	√	√	√	-	-	-	√	-
5.	(Zulfia, 2021)	Implementasi Algoritma <i>Decision Tree</i> Dan <i>K-Nearest Neighbor</i> Pada Klasifikasi Indeks Standar Pencemar Udara (ISPU)	-	√	√	√	-	√	-	-	-

Penelitian ini melakukan perbandingan akurasi dua algoritma klasifikasi yaitu algoritma *Decision Tree* dan *K-Nearest Neighbor* untuk mengetahui hasil akurasi terbaik penelitian ini dilakukan pada proses klasifikasi pada data Indeks Standar Pencemar Udara (ISPU). Objek penelitian kelas pada klasifikasi menggunakan kelas klasifikasi *multiclass*. Klasifikasi dilakukan dengan model *Machine Learning* bahasa pemrograman *Python*.