JURNAL FASILKOM
Volume 14 No. 2 | Agustus: 520-525

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Implementasi Algoritma Random Forest Untuk Klasifikasi Pencemaran Udara di Wilayah Jakarta Berdasarkan Jakarta Open Data

Rahmad Firdaus[1], Husnul Habibie[2], Yoze Rizki[3]
[1,2,3]Teknik Informatika, Ilmu Komputer, Universitas Muhammadiyah Riau
[1]rahmadfirdaus@umri.ac.id, [2]190401164@student.umri.ac.id, [2]yozerizki@umri.ac.id

Abstract

Air pollution is a world problem that is quite concerning in several countries, including one in Jakarta. DKI Jakarta is one of the cities with the highest ranking for the worst air quality in the world. The Random Forest Algorithm is the development of the Classification and Regression Tree (CART) method which can improve the results of accuracy in generating attributes for each node which is done randomly. This study aims to determine the performance of the Random Forest Algorithm for classification in air pollution data for the Jakarta area in 2016-2021 and to obtain the classification results from the Random Forest Algorithm in air pollution classification for the Jakarta area in 2016-2021. So that this research hopefully can be a reference or reference for researchers about the Random Forest algorithm, in the classification of Air Pollution data. Model performance results from the Random Forest algorithm, the data train gets perfect precision, recall, and F1-score values, namely 100% for all classes and the AUC is also 100%, then on the test data the precision values for each class are also very high, namely 99%. and AUC of 99.96%. Classification results from the Random Forest algorithm get an accuracy on the train data of 100% and for the test data get an accuracy on the train data of 99.95%.

Keywords: air pollution, jakarta, random forest, classification, Machine Learning

Abstrak

Pencemaran udara merupakan masalah dunia yang cukup memprihatinkan di beberapa negara, dan termasuk salah satunya di Jakarta. DKI Jakarta merupakan salah satu kota dengan peringkat tertinggi dalam kualitas udara yang terburuk di dunia. Algoritma Random Forest adalah pengembangan dari metode *Classification and Regression Tree* (CART) yang dapat meningkatkan hasil akurasi dalam membangkitkan atribut untuk setiap node yang dilakukan secara acak. Pada penelitian ini bertujuan untuk mengetahui Performa Algoritma *Random Forest* terhadap klasifikasi dalam data pencemaran udara wilayah Jakarta tahun 2016- 2021 dan untuk mendapatkan hasil klasifikasi dari Algorima *Random Forest* dalam klasifikasi pencemaran udara wilayah Jakarta tahun 2016-2021. Sehingga penelitian ini semoga dapat menjadi rujukan atau acuan bagi peneliti tentang algoritma *Random Forest*, dalam klasifikasi data Pencemaran Udara. Hasil performa model dari algoritma Random Forest, pada data train mendapatkan nilai *precision*, *recall*, dan F1-*score* yang sempurna yaitu 100% disemua kelas dan AUC juga sebesar 100%, lalu pada data test pada nilai precision untuk setiap kelas juga sangat tinggi yaitu 99%, dan AUC sebesar 99,96%. Hasil klasifikasi dari algoritma *Random Forest* mendapatkan akurasi pada data train sebesar 100% dan untuk data test mendapatkan akurasi pada data train sebesar 99,95%.

Kata kunci: pencemaran udara, jakarta, random forest, klasifikasi, Pembelajaran Mesin

1. Pendahuluan

Pencemaran udara telah berkembang menjadi salah satu isu lingkungan yang paling sering dikeluhkan oleh masyarakat. Mempertimbangkan dampak signifikan terhadap kesehatan masyarakat, polusi atau pencemaran udara merupakan masalah dunia yang cukup memprihatinkan di beberapa negara, dan termasuk salah satunya di Jakarta, Kota metropolitan ini merupakan salah satu kota dengan tingkat polusi udara yang relatif tinggi.

Pencemaran udara terjadi karena aktivitas dari sumber-sumber yang dapat bergerak maupun yang tidak bergerak, termasuk transportasi, sektor industri, dan kegiatan di rumah tangga. Selain itu, pertumbuhan penduduk dan urbanisasi yang cepat juga berkontribusi secara tidak langsung terhadap masalah pencemaran udara, serta pembangunan daerah yang tinggi, tidak seimbang dan rendahnya kesadaran masyarakat terhadap pencemaran udara [1].

Pencemaran udara terjadi ketika zat-zat atau energi dari berbagai sumber manusia masuk ke dalam udara, menyebabkan penurunan kualitas udara hingga mencapai tingkat di mana udara tidak lagi mampu memenuhi fungsinya dengan baik [2] Hal itu dikarenakan keberadaan sejumlah besar Substansi fisik, biologis, atau kimia di atmosfer bumi yang berpotensi membahayakan kesehatan manusia serta organisme hidup lainnya. Dengan meningkatnya polusi udara, prakiraan kualitas udara dan sistem peringatan dini dapat dikembangkan untuk memantau dan mengendalikan kualitas pencemaran udara.

Berdasarkan *Air Quality Live Index* (AQLI), DKI Jakarta menempati peringkat keenam sebagai kota dengan kualitas udara terburuk pada bulan April 2021. Pada nilai indeks AQLI, tercatat bahwa Jakarta mencatatkan indeks kualitas udara sebesar 156 yang dikategorikan sebagai tidak sehat. Penurunan kualitas udara disebabkan oleh polutan utama PM2.5, yang

P-ISSN : 2089-3353
E-ISSN : 2808-9162

harus tetap berada di Partikel dengan diameter kurang dari 10 mikron yang ada di udara. Di DKI Jakarta, polutan ini tercatat mencapai konsentrasi 57 mikron per meter kubik, yang menunjukkan bahwa kualitas udara di wilayah tersebut sangat buruk [3] Kota DKI Jakarta dipilih berdasarkan hasil pemantauan kualitas udara yang dilakukan oleh United States (US) Air Quality Index (AQI) pada kuartal ketiga tahun 2019 menunjukkan bahwa DKI Jakarta pernah menduduki peringkat pertama sebagai kota dengan kualitas udara terburuk di dunia. Selain itu, selama periode tahun 2017 hingga 2019, kualitas udara di wilayah DKI Jakarta terus memburuk dengan penurunan tingkat kebersihan udara dan peningkatan jumlah hari yang dianggap tidak sehat setiap tahunnya [4].

Tingkat kebersihan udara di suatu wilayah berbeda beda, ada wilayah yang tingkat polusi tinggi, sedang dan ada yang rendah, seperti di wilayah perkotaan yang sebagian memiliki tingkat pencemaran udara yang tinggi dan Sebagian wilayah perdesaan atau hutan yang masih memiliki tingkat polusi yang rendah. Rekomendasi Kualitas Udara WHO 2005 memberikan panduan global tentang nilai ambang dan ambang batas polusi udara yang menjadi faktor penyebab risiko kesehatan. Pedoman tersebut menunjukkan bahwa mengurangi emisi partikel (PM10) dari 20 menjadi 70 mikrogram per meter kubik (ug/m3) dapat mengurangi kematian dan juga tingkat level akibat polusi udara sekitar 15 persen.

Pemerintah Indonesia, melalui Keputusan Badan Pengendalian Dampak Lingkungan (Bapedal) Nomor KEP-107/Kabapedal/11/1997, menetapkan Indeks Standar Pencemar Udara (ISPU) sebagai alat untuk mengevaluasi kualitas udara di suatu daerah dan dampaknya terhadap kesehatan manusia setelah paparan udara selama beberapa jam hingga hari. Semakin tinggi tingkat ISPU, semakin besar potensi risiko kesehatan dari udara yang dihirup. ISPU diklasifikasikan dalam lima kategori, yaitu Baik, Sedang, Tidak Sehat, Sangat Tidak Sehat, dan Berbahaya

Menurut penelitian [4] Penentuan tingkat level ISPU dapat dipermudah melalui penerapan teknik klasifikasi dalam data mining. Data mining adalah metode yang digunakan untuk menggali informasi baru dengan mengidentifikasi aturan atau pola tertentu dari kumpulan data yang besar

Adapun beberapa penelitian sebelumnya yang menggunakan algoritma Klasifikasi Random Forest yaitu pada penelitian Classification Analysis Of Unilak Informatics Engineering Students Using Support Vector Machine (SVM), Iterative Dichotomiser 3 (Id3), Random Forest And K-Nearest Neighbors (KNN) yang mencoba membandingkan algoritma klasifikasi untuk periode studi mahasiswa dengan empat metode, yaitu Support Vector Machine, Iterative Dichotomiser, Random Forest, dan K-Nearest Neighbor. Akhirnya dapat diambil kesimpulan dari penelitian ini, bahwa masing-masing dari algoritma Support Vector Machine, Iterative Dichotomiser 3, Random Forest, dan K-Nearest Neighbor mampu mengklasifikasikan data yang ada, dan didapatkan pada penelitian ini bahwa algoritma Random Forest adalah algoritma dengan tingkat akurasi terbaik dibandingkan dengan tiga algoritma lainnya, dengan persentase akurasi data mencapai 100%. Diikuti oleh algoritma Support Vector Machine dengan akurasi 94,4%, dan Iterative Dichotomiser serta K-Nearest Neighbor dengan nilai persentase sebesar 90% masing-masing [5].

Penelitian lain dilakukan oleh [6] pada penelitian ini dilakukan pengukuran untuk melihat tingkat keparahan penyakit pada daun apel menggunakan metode klasifikasi Random Forest. Menunjukkan bahwa tingkat keparahan pada daun apel menghasilkan tingkat akurasi pada proses pelatihan sebesar 100% dan pengujian sebesar 75.3191%.

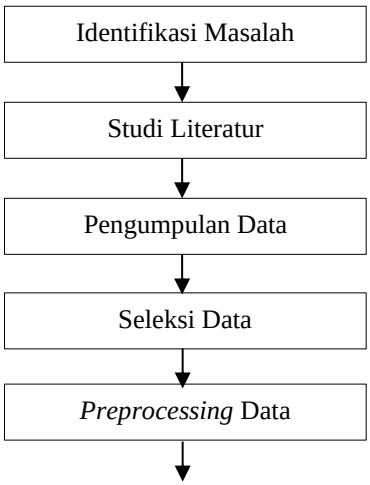
Penelitian lain oleh [7] tujuan penelitian ini adalah untuk mengukur akurasi metode Random Forest dalam memprediksi varian minuman kopi di kedai Konijiwa Banteng yang diminati oleh pelanggan. Menghasilkan tingkat akurasi sebesar 94,12%.

Penelitian lain dilakukan oleh [8] penelitian ini membandingkan performa dari 3 algoritma yaitu Random Forest, Decision Tree dan Support Vectore Machine (SVM). Algoritma Random Forest menghasilkan algoritma terbaik dengan hasil akurasi sebesar 0.7468, Decision Tree sebesar 0.7031 dan Support Vectore Machine (SVM) sebesar 0.65.

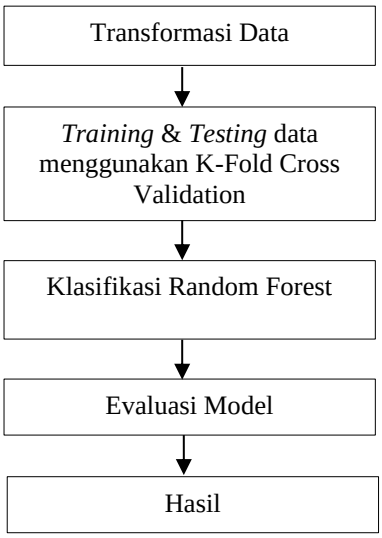
Berdasarkan pemaparan yang sudah dilakukan, tujuan penelitian yang dilakukan adalah untuk mengukur performa Algoritma Random Forest dalam melakukan klasifikasi Pencemaran Udara Di Wilayah Jakarta menggunakan dataset Jakarta Open Data.

2. Metode Penelitian

Metodologi penelitian adalah panduan atau langkah-langkah sistematis dalam pelaksanaan penelitian. Metodologi ini diterapkan dengan tujuan untuk memperoleh hasil yang sesuai dengan yang diharapkan



P-ISSN : 2089-3353
E-ISSN : 2808-9162



Gambar 1. Metode Penelitian

2.1. Studi Literatur

Tahapan studi literatur adalah proses yang melibatkan pengumpulan, penilaian, dan analisis berbagai sumber informasi yang relevan dengan topik penelitian. Tujuan dari tahapan ini adalah untuk memahami konteks dan landasan teori dari penelitian yang akan dilakukan.

Tahapan dalam merumuskan penelitian ini mencakup pengumpulan informasi yang relevan dengan topik penelitian, yaitu berkaitan dengan kualitas udara, ISPU, dan berbagai algoritma klasifikasi. Informasi tersebut diperoleh dari berbagai sumber, termasuk buku, artikel, jurnal, dan dokumen lainnya.

2.2. Persiapan Dataset

Persiapan dataset adalah langkah-langkah awal dalam proses analisis data yang melibatkan pemrosesan dan pembersihan data sebelum analisis lebih lanjut dilakukan. Tujuan dari persiapan dataset adalah untuk memastikan bahwa data dalam kondisi yang baik dan siap untuk digunakan dalam model atau analisis.

Dataset yang bersumber dari Jakarta Open Data berjumlah 10955 baris dan 8 kolom. Dataset mencakup Indeks Standar Pencemar Udara (ISPU) yang diukur dari lima stasiun pemantau kualitas udara (SPKU) yang tersebar di Provinsi DKI Jakarta mulai periode Januari 2016 hingga Desember 2021 dengan berkas berkas yang berupa file CSV yang di unduh satu per satu di website Jakarta Open data. Dataset tersebut telah dipublikasi dan didokumentasi pada website Jakarta Open Data. Data ini diperoleh dari Badan Pengendalian Lingkungan Hidup (BPLHD) Provinsi DKI Jakarta. Data meliputi informasi tingkat polusi udara (seperti PM10, PM2.5, SO2, NO2, dan O3), waktu pengukuran, koordinat geografis, serta informasi cuaca (seperti suhu, kelembaban, dan kecepatan angin) pada setiap titik pengukuran.

2.3. Data *Selection*

Dataset Indeks Standar Pencemar Udara (ISPU) DKI Jakarta dikumpulkan dan diseleksi dari situs web pemerintah DKI Jakarta, data.jakarta.go.id, dalam format CSV. Data ini mencakup informasi dari masing-masing Stasiun Pemantau Kualitas Udara (SPKU) di wilayah Jakarta, yang diperoleh dari Januari 2016 hingga Desember 2021

Dataset mencakup Indeks Standar Pencemar Udara (ISPU) yang diukur dari lima stasiun pemantau kualitas udara (SPKU) di Provinsi DKI Jakarta. Data yang diperoleh memiliki atribut dan label, namun belum teratur atau bersih

2.4. *Preprocessing* Data

Preprocessing data adalah tahap awal dalam *pipeline* analisis data atau *machine learning* yang melibatkan serangkaian langkah untuk menyiapkan data mentah agar siap untuk analisis atau pemodelan. Tujuan dari *preprocessing* data adalah untuk memastikan bahwa data yang digunakan dalam model atau analisis dalam kondisi optimal dan dapat memberikan hasil yang akurat Pada tahap ini akan dilakukan proses pembersihan data dan menghilangkan *missing value*.

2.5. Transformasi Data

Transformasi data adalah proses mengubah data dari format atau struktur aslinya menjadi format atau struktur yang lebih sesuai untuk analisis atau pemodelan. Transformasi ini dilakukan untuk memudahkan proses analisis, memperbaiki kualitas data, atau memenuhi kebutuhan khusus dari algoritma *machine learning*.

Pada tahap ini, dataset yang masih terpisah dalam berkas CSV diintegrasikan dan atribut yang ada disesuaikan, sehingga membentuk kesatuan dataset yang terstruktur. Hal ini bertujuan untuk mempermudah proses analisis selanjutnya

2.6. K-Fold Cross Validation

Untuk memahami sejauh mana model mampu mengklasifikasikan data dengan akurat dengan membagi data menggunakan K-Fold Cross-Validation. Tahap pembagian data menjadi 5 fold dengan menggunakan K-Fold Cross- Validation.

2.7. Evaluasi Model

Pada tahap ini dilakukan uji validasi untuk mengevaluasi hasil dengan menggunakan nilai akurasi dari algoritma Random Forest menggunakan metode *Confusion Matrix*. Confusion matrix adalah alat evaluasi yang digunakan untuk menilai kinerja model klasifikasi dengan membandingkan hasil prediksi model dengan label yang sebenarnya. Confusion matrix menyediakan informasi rinci tentang bagaimana model melakukan klasifikasi dengan mengklasifikasikan data ke dalam kategori yang benar dan salah.

P-ISSN : 2089-3353
E-ISSN : 2808-9162

3. Hasil dan Pembahasan

Pada bab ini akan diuraikan bagaimana tahapan penelitian yang telah dilakukan:

3.1. Persiapan Dataset

Data yang digunakan merupakan data pencemaran udara di wilayah Jakarta yang dikumpulkan dari website Jakarta open data. File yang berada di folder terpisah digabung menggunakan modul glob yang di impor dari Google Drive untuk di olah di Google Collab:

```
import glob

folder_path = "/content/drive/MyDrive/SKRIPSI/Datasets/Jadi"
file_list = glob.glob(f"{folder_path}/*/*")  # Menggunakan pola
pencarian *.extensi

for file_path in file_list:
    print(file_path)
```

Gambar 2. Persiapan Dataset

3.2. Data *Selection*

Pada tahapan ini, akan dilakukan pemilihan atribut. Dengan cara *combine* data, sehingga dapat memilih dan menampilkan hanya atribut-atribut tertentu dari dataset yang ada dalam *DataFrame combine_data*.

Untuk hasil dari tahapan ini berhasil menyeleksi atribut dan membersihkan data dari atribut dan data yang tidak diperlukan atau data kosong. Berikut ini hasil dari proses tahapan data *selection*:

	pm10	so2	co	o3	no2	max	critical	categori
0	30	20	10	32	9	32	O3	BAIK
1	27	22	12	29	8	29	O3	BAIK
2	39	22	14	32	10	39	PM10	BAIK
3	34	22	14	38	10	38	O3	BAIK
4	35	22	12	31	9	35	PM10	BAIK
...	...	...	...	...	...	..	...	...
10990	35	11	---	36	9	36	O3	BAIK
10991	48	11	---	66	15	66	O3	SEDANG
10992	---	---	---	---	---	0	NaN	TIDAK ADA DATA
10993	---	---	---	---	---	0	NaN	TIDAK ADA DATA
10994	---	---	---	---	---	0	NaN	TIDAK ADA DATA

[10995 rows x 8 columns]

Gambar 3. Pemilihan Atribut

3.3. *Data Preprocessing*

Proses yang dilakukan pada tahapan *preprocessing* yaitu *data cleaning*. Pada tahap ini akan dilakukan proses pembersihan data dari data yang tidak diperlukan dan data yang hilang, dengan cara melakukan beberapa pemrosesan pada DataFrame 'df', yang telah dipilih pada tahapan sebelumnya dengan memilih kolom-kolom tertentu dari DataFrame combined_data.

```
df = df[df.pm10 != "---"]
df = df[df.so2 != "---"]
df = df[df.o3 != "---"]
df = df[df.co != "---"]
df = df[df.no2 != "---"]
df = df[df.critical != "---"]
df = df[df.categori != "---"]
df.dropna(subset = ["critical"], inplace=True)
df.dropna(subset = ["max"], inplace=True)
df= df[df['max'] != 0]
```

Gambar 4. *Cleaning data*

Setelah melakukan semua pemrosesan tersebut, DataFrame "df" akan berisi data yang telah diolah dan terbebas dari nilai yang tidak relevan seperti '---', nilai "NaN" (*missing value*), dan nilai 0 pada kolom 'max'.

3.4. Pengecekan *missing value*

Setelah beberapa kode diatas dijalankan, selanjutnya untuk pengecekan data yang kosong atau yang tidak di perlukan.

	pm10	so2	co	o3	no2	max	critical	categori
0	30	20	10	32	9	32	1.0	1.0
1	27	22	12	29	8	29	1.0	1.0
2	39	22	14	32	10	39	2.0	1.0
3	34	22	14	38	10	38	1.0	1.0
4	35	22	12	31	9	35	2.0	1.0
...	...	...	...	...	...	...	...	...
10974	66	10	35	46	11	66	2.0	2.0
10975	59	11	33	29	20	59	2.0	2.0
10976	65	11	36	32	14	65	2.0	2.0
10982	56	13	12	52	12	56	2.0	2.0
10984	44	13	20	26	13	44	2.0	1.0

9040 rows × 8 columns

Gambar 5. Pengecekan *Missing Value*

Berdasarkan output tersebut, data telah berhasil di bersihkan dan menyusut menjadi 9040 baris data dan 8 kolom.

3.5. Transformasi Data

Tahapan ini dilakukan untuk dapat membantu memastikan bahwa data yang dimasukkan kedalam model memiliki format yang sesuai dan siap digunakan untuk pelatihan. Dataset akan dilakukan pemetaan atau penggantian nilai pada suatu kolom berdasarkan library yang diberikan menggunakan metode `map()`.

Pada bagian pertama, menggunakan map() pada kolom 'categori' untuk mengganti nilai-nilai kategori menjadi angka sebagai berikut: 'BAIK' diganti menjadi 1, 'SEDANG' diganti menjadi 2, 'TIDAK SEHAT' diganti menjadi 3, 'SANGAT TIDAK SEHAT' diganti menjadi 4, 'BERBAHAYA' diganti menjadi 5.

Pada bagian kedua, menggunakan map() pada kolom 'critical' untuk mengganti nilai-nilai kategori menjadi angka sebagai berikut: 'O3' diganti menjadi 1, 'PM10'

P-ISSN : 2089-3353
E-ISSN : 2808-9162

diganti menjadi 2, 'SO2' diganti menjadi 3, 'PM25' diganti menjadi 4, 'CO' diganti menjadi 5

Dengan menggunakan metode map(), dapat dengan mudah melakukan penggantian nilai berdasarkan kamus yang telah ditentukan. Setelah proses map() selesai, nilai- nilai pada kolom 'categori' dan 'critical' akan berubah menjadi angka sesuai dengan pemetaan yang telah diberikan. Berikut hasilnya :

	pm10	so2	co	o3	no2	max	critical	categori
0	30	20	10	32	9	32	1.0	1.0
1	27	22	12	29	8	29	1.0	1.0
2	39	22	14	32	10	39	2.0	1.0
3	34	22	14	38	10	38	1.0	1.0
4	35	22	12	31	9	35	2.0	1.0
...	...	...	...	...	...	...	...	...
566	111	31	25	58	10	111	2.0	3.0
567	66	27	13	51	7	66	2.0	2.0
569	64	26	25	53	9	64	2.0	2.0
572	59	20	14	60	9	60	1.0	2.0
574	80	22	26	44	9	80	2.0	2.0

500 rows × 8 columns

Gambar 6. Transformasi Data

Pada bagian atribut “critical” dan label “categori” telah berganti nilai menjadi angka.

3.6. Pemodelan

Untuk memahami sejauh mana model mampu mengklasifikasikan data dengan akurat dengan membagi data menggunakan *K-Fold Cross-Validation*.

Tahapan pertama merupakan proses pembagian data menjadi 5 fold dengan menggunakan *K-Fold Cross-Validation.*

```
Fold:1, Train set: 7232, Test set:1808
Fold:2, Train set: 7232, Test set:1808
Fold:3, Train set: 7232, Test set:1808
Fold:4, Train set: 7232, Test set:1808
Fold:5, Train set: 7232, Test set:1808
```

Gambar 7. Lima Fold Pembagian Data

Diperoleh hasil score dari setiap “fold” dengan rata rata nilai 99,93%.

```
Score setiap fold:
[0.99889381         1.     0.99834071     1.     0.9994469 ]
Average score: 99.93%
```

Gambar 8. Hasil Rata-Rata Setiap Fold

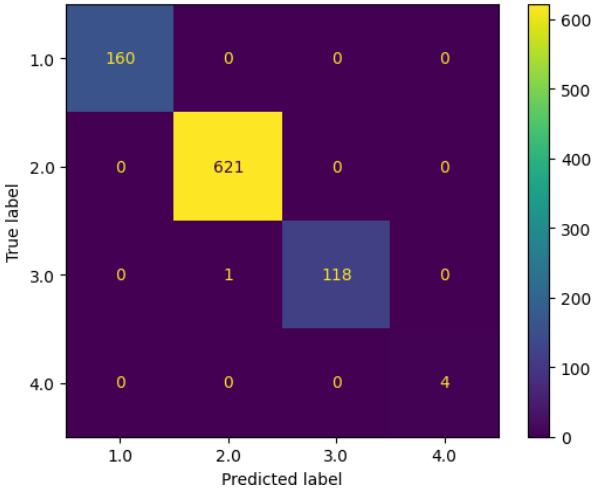
Tahapan kedua merupakan proses klasifikasi data menggunakan algoritma Random Forest. Dari proses yang telah dilakukan mendapatkan nilai akurasi dari data test, dan data train yang menghasilkan akurasi 0,9994 untuk data test, dan 1,000 atau 1,0 untuk data train.

```
Nilai akurasi pada data test  : 0.9994
Nilai akurasi pada data train : 1.0000
```

Gambar 9. Hasil Pemodelan

3.7. Evaluasi Model (*Confusion Matrix*)

Hasil Confusion Matrix yang mencantumkan jumlah prediksi yang benar dan salah untuk setiap kelas, serta laporan klasifikasi yang menyediakan beberapa metrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk setiap kelas.



Gambar 10. Hasil Confusion Matrix

3.8. Hasil Penelitian

```
Train Result:
================================================
Accuracy Score: 100.00%
------------------------------------------------
CLASSIFICATION REPORT:
                1.0     2.0    3.0   4.0  accuracy  macro avg  weighted avg
precision       1.0     1.0    1.0   1.0       1.0        1.0           1.0
recall          1.0     1.0    1.0   1.0       1.0        1.0           1.0
f1-score        1.0     1.0    1.0   1.0       1.0        1.0           1.0
support      1286.0  4964.0  957.0  25.0       1.0     7232.0        7232.0
------------------------------------------------
Confusion Matrix:
 [[1286    0    0    0]
 [   0 4964    0    0]
 [   0    0  957    0]
 [   0    0    0   25]]

Test Result:
================================================
Accuracy Score: 99.94%
------------------------------------------------
CLASSIFICATION REPORT:
                1.0          2.0          3.0  4.0  accuracy    macro avg  \
precision       1.0     0.999195     1.000000  1.0  0.999447     0.999799
recall          1.0     1.000000     0.995816  1.0  0.999447     0.998954
f1-score        1.0     0.999597     0.997904  1.0  0.999447     0.999375
support       321.0  1241.000000  239.000000  7.0  0.999447  1808.000000

              weighted avg
precision         0.999447
recall            0.999447
f1-score          0.999446
support        1808.000000
------------------------------------------------
Confusion Matrix:
 [[ 321    0    0    0]
 [   0 1241    0    0]
 [   0    1  238    0]
 [   0    0    0    7]]
```

Gambar 11. Hasil Penelitian

Berikut hasil yang menambilkan indikator, seperti nilai AUC, lalu ada akurasi, *precision*, *recall*, dan f1-*score*. Hasil performa model dari algoritma *Random Forest*,

P-ISSN : 2089-3353
E-ISSN : 2808-9162

pada data train mendapatkan nilai precision, recall, dan F1-score yang sempurna yaitu 100% disemua kelas dan AUC juga sebesar 100%, lalu pada data test pada nilai *precision* untuk setiap kelas juga sangat tinggi yaitu 99%, dan AUC sebesar 99,96%. Hasil klasifikasi dari algoritma Random Forest mendapatkan akurasi pada data *train* sebesar 100% dan untuk data *test* mendapatkan akurasi pada data train sebesar 99,95%.

4. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah diuraikan pada bab sebelumnya, dapat diambil kesimpulan bahwa:

1. Hasil dari evaluasi dan validasi, diketahui bahwa *Random Forest*, pada data train mendapatkan nilai *precision*, *recall*, dan F1-*score* yang sempurna yaitu 100% disemua kelas dan AUC juga sebesar 100%, lalu pada data test pada nilai *precision* untuk setiap kelas juga sangat tinggi yaitu 99%, dan AUC sebesar 99,96%.

2. Hasil klasifikasi dari algoritma Random Forest mendapatkan akurasi pada data *train* sebesar 100% dan untuk data test mendapatkan akurasi pada data train sebesar 99,95%. Dari hasil tersebut dapat menjadi bukti bahwa algoritma *Random Forest* dapat digunakan untuk klasifikasi data pencemaran udara dan model Random Forest yang telah dilatih memiliki kinerja yang sangat baik dalam mengklasifikasikan data, baik pada data train maupun data test. Hal ini ditunjukkan oleh akurasi yang tinggi dan nilai presisi, *recall*, dan F1-*score* yang baik untuk semua kelas.

Daftar Rujukan

[1] A. H. R. Inaku and C. Novianus, "Pengaruh Pencemaran Udara PM 2,5 dan PM 10 Terhadap Keluhan Pernapasan Anak di Ruang Terbuka Anak di DKI Jakarta," *ARKESMAS (Arsip Kesehat. Masyarakat)*, vol. 5, no. 2, pp. 9–16, 2020, doi: 10.22236/arkesmas.v5i2.4990.

[2] P. P. R. Indonesia, "Peraturan Pemerintah Republik Indonesia Nomor 41 Tahun 1999 Tentang Pengendalian Pencemaran Udara." https://ppkl.menlhk.go.id/%0A.

[3] I. N. I. Adinda Amalia, Ati Zaidiah, "PREDIKSI KUALITAS UDARA MENGGUNAKAN ALGORITMA K- NEAREST NEIGHBOR," *J. Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 2, pp. 496–507, 2022, doi: 10.33387/jiko.v4i2.2871.

[4] S. S. A. Umri *et al.*, "Analysis and Comparison of Classification Algorithm in Air," *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 2, pp. 98–104, 2021, doi: 10.33387/jiko.

[5] H. Sunaryanto, M. A. Hasan, and G. Guntoro, "Classification Analysis of Unilak Informatics Engineering Students Using Support Vector Machine (SVM), Iterative Dichotomiser 3 (ID3), Random Forest and K-Nearest Neighbors (KNN)," *IT J. Res. Dev.*, vol. 7, no. 1, pp. 36–42, 2022, doi: 10.25299/itjrd.2022.8912.

[6] L. Ratnawati and D. R. Sulistyaningrum, "Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel," *J. Sains dan Seni ITS*, vol. 8, no. 2, 2020, doi: 10.12962/j23373520.v8i2.48517.

[7] Suci Amaliah, M. Nusrang, and A. Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 3, pp. 121–127, 2022, doi: 10.35580/variansiunm31.

[8] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," *E-Bisnis J. Ilm. Ekon. dan Bisnis*, vol. 13, no. 2, pp. 67–75, 2020, doi: 10.51903/e-bisnis.v13i2.247.