

# KLASIFIKASI TINGKAT PENCEMARAN UDARA PADA SEKTOR INDUSTRI DENGAN METODE *RANDOM FOREST*

CLASSIFICATION OF AIR POLLUTION IN INDUSTRIAL SECTOR USING  
RANDOM FOREST

**Suci Cahaya Hati Nasution<sup>1</sup>, Fibri Rakhmawati<sup>2</sup>, Riri Syafitri Lubis<sup>3</sup>**

<sup>1,2,3</sup>Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri  
Sumatera Utara

<sup>1,2,3</sup>Universitas Islam Negeri Sumatera Utara; Jl. William Iskandar Ps. V Medan 20371, (+6261)  
6615683, 6622925, Affiliation ID : 60110865

e-mail : \*<sup>1</sup>[sucicahaya242@gmail.com](mailto:sucicahaya242@gmail.com), <sup>2</sup>[Fibri\\_rakhmawati@uinsu.ac.id](mailto:Fibri_rakhmawati@uinsu.ac.id),  
<sup>3</sup>[riri\\_syafitri@uinsu.ac.id](mailto:riri_syafitri@uinsu.ac.id)

## **Abstrak**

*Industri merupakan salah satu penggerak perekonomian Indonesia. Kegiatan industri di Indonesia di proyeksikan pada tahun 2020 menjadi 13 kali lebih besar dari tahun 1990. Akibat aktifitas industri udara seringkali menurun kualitasnya. Pencemaran udara dirasakan terus meningkat yang menyebabkan penurunan kualitas lingkungan yang berdampak pada masalah kesehatan. Penelitian bertujuan untuk mengetahui tingkat akurasi dan pengaruh variabel dalam klasifikasi tingkat pencemaran udara pada sektor industri. Data yang digunakan merupakan data sekunder yang diambil dari Badan Lingkungan Hidup kota Medan tahun 2018 sebanyak 20 perusahaan. Metode yang digunakan adalah metode random forest. Metode ini merupakan salah satu metode ensemble yang bertujuan untuk meningkatkan hasil akurasi dari suatu klasifikasi. Klasifikasi pada penelitian ini dibagi menjadi tiga kelas yaitu rendah, sedang, dan tinggi, dengan jumlah pohon sebanyak 500 pohon. Tingkat akurasi berdasarkan pemodelan random forest diperoleh sebesar 95% prediksi model dinyatakan benar. Variabel yang memiliki pengaruh tertinggi yaitu sulfur dioksida dan nitrogen dioksida dan variabel terendah adalah opasitas dan partikulat.*

**Kata kunci :** industri, pencemaran udara, *random forest*.

## **Abstract**

*Industry is one of boosters for the Indonesian economy. In 2020, Indonesian industrial work is projected 13 times greater than in 1990. Industrial works result air quality decrease. Air pollution is perceived to increase, so it leads environmental quality decrease causes impact on health problems. The research aims to determine the accuracy level and influence of variables in the classification of air pollution levels in the industrial sector. 20 manufactories is taken from Medan Environment Agency in 2018 and used as secondary data for this research. The research used random forest method, one of ensemble methods that aims to improve the accuracy of a classification. This research divided the classification into three classes. Random forest modeling obtained 95% accuracy rate which proved model prediction correct. The highest influential variables were sulfur dioxide and nitrogen dioxide and followed by opacity and particulate matter (dust) as the lowest influential variables.*

**Keywords :** industry, air pollution, *random forest*.

## 1. PENDAHULUAN

Industri merupakan salah satu penggerak perekonomian negara Indonesia, dengan zaman yang semakin berkembang mengakibatkan sektor industri juga semakin meningkat dengan kemajuan teknologi yang semakin maju serta kebutuhan yang semakin meningkat. Industri di Indonesia juga meningkat dengan pesat pada tahun 1990–2020, kegiatan industri di Indonesia diproyeksikan pada tahun 2020 menjadi 13 kali lebih besar dari yang terdapat pada tahun 1990<sup>1</sup>.

Tidak dapat dipungkiri bahwa pembangunan pada sektor industri memberi dampak positif dan dampak negatif. Adapun beberapa dampak positif dari industri adalah membuka lapangan pekerjaan, menambah pemasukan pajak daerah, stabilitas ekonomi, meningkatkan pemasukan devisa bagi negara, mengurangi pengangguran. Sedangkan salah satu dampak negatifnya adalah meningkatkan pencemaran lingkungan antara lain pencemaran air, pencemaran tanah, dan pencemaran udara.

Menurut peraturan pemerintah Republik Indonesia No.41 Tahun 1999 tentang pengendalian pencemaran udara, pencemaran udara adalah masuknya atau dimasukkannya zat, energi dan/atau komponen lain ke dalam udara ambien oleh kegiatan manusia, sehingga mutu udara ambien turun sampai tingkat tertentu yang menyebabkan udara ambien tidak memenuhi fungsinya.<sup>2</sup>

Udara dimana didalamnya terkandung sejumlah oksigen (O), yang merupakan komponen esensial bagi kehidupan, baik manusia maupun makhluk hidup lainnya. Udara merupakan campuran dari berbagai gas dengan konsentrasi yang berbeda-beda. Gas-gas tersebut diantaranya adalah nitrogen dalam bentuk N<sub>2</sub> sebanyak 78%, oksigen dalam bentuk O<sub>2</sub> sebanyak 21%, sementara argon (Ar) hanya 1% dari total gas. Gas-gas karbondioksida (CO<sub>2</sub>), helium (He), neon (ne), Xenon (Xe), dan Krypton (Kr) masing-masing hanya terdapat 0,01% dari total gas. Selain itu, ada beberapa jenis gas dalam jumlah sedikit seperti Metana (CH<sub>4</sub>), karbon monoksida (CO), amoniak (NH<sub>3</sub>), dinitrogen (N<sub>2</sub>O), dan Hidrogen Sulfida (H<sub>2</sub>S) dan apabila gas-gas tersebut dalam jumlah yang besar gas-gas tersebut berpotensi menyebabkan pencemaran udara.<sup>3</sup> Udara dikatakan “normal” dan dapat mendukung kehidupan manusia apabila komposisinya seperti yang disebutkan di atas. Sedangkan apabila terjadi penambahan gas-gas lain yang menimbulkan gangguan serta perubahan komposisi tersebut, maka dikatakan udara sudah tercemar atau terpolusi.

Akibat aktifitas industri udara seringkali menurun kualitasnya. Perubahan kualitas ini dapat berupa perubahan sifat-sifat fisik maupun kimiawi, dapat berupa pengurangan maupun penambahan salah satu komponen kimia yang terkandung dalam udara. Kualitas udara dipergunakan untuk kehidupan tergantung dari lingkungannya. Kemungkinan disuatu tempat dijumpai debu yang bertebaran dimana-mana dan berbahaya bagi kesehatan. Demikian juga suatu kota yang terpolusi oleh asap kendaraan bermotor atau angkutan yang dapat menimbulkan gangguan kesehatan.<sup>4</sup>

Pencemaran udara dirasakan semakin hari semakin meningkat, terutama pada lokasi industri yang kurang memperhatikan dampak buruk dari aktifitas industri berlebih pada lingkungannya. Pencemaran udara dapat menurunkan kualitas lingkungan yang berdampak pada masalah kesehatan. Salah satu sumber yang menyebabkan pencemaran udara adalah industri, dimana industri-industri yang aktif mengeluarkan gas emisi hasil dari pengolahan suatu industri seperti SO<sub>2</sub>, NO<sub>2</sub>, HC, TSP, NH<sub>3</sub>, O<sub>3</sub>, dan lain-lain.

Karena pencemaran udara memiliki dampak yang serius pada kehidupan sehari-hari, dan bagaimana cara yang tepat untuk mengklasifikasikan pencemaran udara merupakan salah satu tujuan penelitian ini, maka peneliti bertujuan menggunakan metode *Random Forest* untuk menentukan klasifikasi tingkat pencemaran udara pada sektor industri di kota medan.

Metode *Random Forest* merupakan pengembangan dari metode *Classification and Regression Tree* (CART) dengan menerapkan metode *bootstrap aggregating* dan *random feature selection*. Kelebihan metode ini antara lain dapat menghasilkan akurasi yang lebih tinggi, dapat mengatasi data jumlah yang besar secara efisien, dan tidak terdapat pemangkasan

variabel seperti pada algoritma pohon klasifikasi tunggal, selain menghasilkan akurasi hasil prediksi yang sangat tinggi, *Random Forest* juga menghasilkan nilai tingkat kepentingan variabel prediktor dalam pengklasifikasian variabel respon.<sup>5</sup>

Klasifikasi adalah tahap terakhir pada desain sistem ini. Tingkat pencemaran udara pada sektor industri akan diklasifikasikan berdasarkan tiga tingkat keparahan, yaitu rendah, sedang, dan tinggi.

Berdasarkan latar belakang diatas, maka dapat dirumuskan permasalahan sebagai tentang tingkat akurasi klasifikasi tingkat pencemaran udara pada sektor industri di kota Medan dengan metode *Random Forest* dan tentang tingkat pengaruh sulfur dioksida, nitrogen dioksida, opasitas, dan partikulat dalam klasifikasi tingkat pencemaran udara pada sektor industri di kota Medan dengan metode *Random Forest*.

Adapun tujuan dari penelitian ini adalah untuk mengetahui tingkat akurasi klasifikasi tingkat pencemaran udara pada sektor industri di kota Medan dengan metode *Random Forest* dan tingkat pengaruh sulfur dioksida, nitrogen dioksida, opasitas, dan partikulat dalam klasifikasi tingkat pencemaran udara pada sektor industri di kota Medan dengan metode *Random Forest*.

Manfaat dari pembahasan masalah ini adalah dapat memberikan gambaran dan wawasan tentang klasifikasi tingkat pencemaran udara pada sektor industri di kota Medan menggunakan metode *Random Forest*, dan untuk mengetahui klasifikasi tingkat pencemaran udara pada sektor industri di kota medan melalui disiplin ilmu bidang Matematika.

## 2. METODE PENELITIAN

Penelitian ini dilaksanakan di Dinas Lingkungan Hidup Kota Medan di jalan Jendral Besar A.H. Nasution No. 32, Pangkalan Mahsyur, Kec. Medan Johor, Kota Medan.

Jenis data yang digunakan pada penelitian ini adalah data kuantitatif. Data kuantitatif adalah data yang dapat diukur atau dihitung secara langsung sebagai variabel angka bilangan. Dalam penelitian ini bersumber dari data sekunder yang diperoleh dari Badan Lingkungan Hidup Kota Medan. Data tersebut berupa data pemantauan kualitas emisi sumber tidak bergerak pada tahun 2018 dari berbagai industri yang ada di kota Medan.

Variabel yang digunakan dalam penelitian ini adalah variabel dependen (respon) dan variabel independen (prediktor). Variabel dependen yang digunakan adalah hasil emisi pembakaran industri yang menyebabkan pencemaran udara. Sedangkan variabel independen yang digunakan adalah industri-industri yang ada di kota Medan. Adapun variabel independen dan dependen akan dijelaskan dalam tabel berikut ini :

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Kategori
Y	Jenis Industri	Rendah, sedang, atau tinggi
X <sub>1</sub>	Sulfur Dioksida	-
X <sub>2</sub>	Nitrogen Dioksida	-
X <sub>3</sub>	Partikulat	-
X <sub>4</sub>	Opasitas	-

Berikut ini adalah penjelasan tentang masing-masing variabel independent dan dependen pada penelitian ini :

1. Jenis industri adalah seluruh industri yang ada di kota Medan. Sebelum menggunakan metode *random forest* masing-masing industri di kota Medan sudah di *cluster* terlebih dahulu untuk pelabelan kelas papa setiap industri.
2. Sulfur dioksida atau belerang dioksida atau dengan rumus kimia  $SO_2$  adalah gas beracun dengan bau menyengat yang biasanya dihasilkan oleh pemrosesan industri atau gunung berapi.
3. Nitrogen dioksida adalah senyawa kimia yang digunakan sebagai bahan sintesis untuk pembuatan asam nitrit, produksi nitrogen dioksida mencapai jutaan ton setiap tahunnya, senyawa kimia ini merupakan salah satu polutan udara utama. Nitrogen dioksida merupakan gas beracun dengan bau menyengat gas ini berwarna merah kecoklatan.
4. Partikulat atau debu adalah partikel yang dihasilkan oleh proses mekanis seperti pada industri-industri besi dan baja. Debu merupakan partikulat yang berukuran antara 1 mikron sampai dengan 100 mikron.
5. Opasitas atau disebut juga dengan tingkat kepekatan asap. Asap merupakan perpaduan atau campuran karbon dioksida, air, zat yang terdifusi di udara, zat partikulat, hidrokarbon, zat kimia organik, nitrogen oksida, dan mineral. Tingkat kepekatan asap dan komposisi asap tergantung dari beberapa faktor yaitu jenis bahan bakar, kelembapan, temperatur api, dan kondisi angin.

Sampel yang digunakan adalah sebanyak 20 industri di kota Medan dan tingkat pencemaran udara pada sektor industri akan diklasifikasikan berdasarkan tiga tingkat keparahan, yaitu rendah, sedang, dan tinggi.

Penelitian ini menggunakan metode *random forest* yang akan digunakan untuk mengklasifikasi tingkat pencemaran udara pada sektor industri di kota Medan. Perangkat lunak yang digunakan untuk mendukung proses penelitian ini adalah R Studio.

Adapun urutan dari analisis datanya adalah sebagai berikut :

1. Instal *Packages*  
*Packages* adalah kumpulan fungsi, data, dan kode. Lakukan penginstalan *packages* yang dibutuhkan dalam analisis terlebih dahulu agar fungsi, data dan kode dapat terbaca oleh perangkat lunak R. Adapun *packages* yang digunakan adalah sebagai berikut :

Tabel 1 Packages

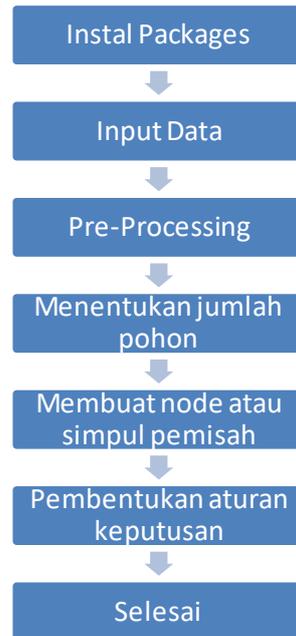
<i>Packages</i>	Fungsi
Readxl	Import data format excel
randomForest	Pemodelan klasifikasi <i>random forest</i>
Devtools	Menginstal reptree
Reptree	Plot pohon

2. Input data  
Data sudah dipersiapkan dalam bentuk *microsoft excel* untuk kebutuhan pengolahan data agar dapat dibaca oleh software *R Studio*.
3. *pre-processing*  
Pada tahap *pre-processing* data berbentuk *microsoft excel* tadi memasuki tahapan *pre-processing* yang didalamnya merupakan persiapan untuk kebutuhan *machine learning*. Yaitu menentukan *feature* (X) dan *label* (Y), dan melakukan *label encoder*.
4. Menentukan jumlah pohon  
Pada tahap ini dilakukan penentuan jumlah k pohon yang akan dibentuk untuk mendapatkan hasil akurasi terbaik. Kemudian membentuk pohon.
5. Membuat *node* atau simpul pemisah, *node* atau simpul akar pohon *n* dengan mengambil sampel secara *random*

Dilakukan *bagging* atau pengambilan data sampel secara *random* untuk membuat *tree* dengan memasukkan data tersebut kedalam perhitungan *gini* guna menemukan *node* atau simpul teratas dan *node* atau simpul pemisah masing-masing *tree*.

6. Pembentukan aturan keputusan  
Setelah proses berlangsung dan *tree* telah terbentuk maka terbentuklah aturan keputusan yang merupakan hasil model prediksi yang akan dijadikan model perhitungan pada sistem.
7. Selesai

Berikut ini adalah diagram alur keputusan pengolahan data :



Gambar 1 Diagram Alur Keputusan

### 3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam analisis ini adalah hasil analisis terhadap kandungan parameter sulfur dioksida, nitrogen dioksida, opasitas, dan partikulat dengan 20 observasi data. Data ini merupakan laporan pemantauan kualitas emisi sumber tidak bergerak terhadap pelaku usaha dan atau kegiatan pada tahun anggaran 2018 yang berasal dari Badan Lingkungan Hidup (BLH) kota Medan. Dari data tersebut akan dilakukan klasifikasi dengan metode *Random Forest*, label dari klasifikasi ini adalah tingkat pencemaran udara pada sektor industri di kota Medan dengan kelas ringan, sedang, dan tinggi. Dikarenakan data yang diperoleh belum memiliki kelas maka sebelumnya dilakukan analisis *clustering* untuk mengelompokkan data ke dalam tiga tingkat pencemaran sesuai dengan karakteristik setiap *cluster*.

Setelah dilakukan *clustering* maka di peroleh kelas untuk setiap data sehingga dapat dilakukan analisis klasifikasi *random forest*. Pemodelan ini dilakukan dengan pemodelan *random forest* dibawah ini.

#### **Kode Pemodelan Random Forest**

```
## klasifikasi Random Forest
library(randomForest)
require(caTools)
```

```
data.rf = read_excel("Data.xlsx", sheet="klasifikasi2")
```

```

data.rf1 = data.rf[,5:9]
data.rf1$kelas = factor(data.rf1$kelas, levels =
c("Ringan", "Sedang", "Tinggi"))
rf <- randomForest(Kelas ~ ., data = data.rf1, mtry=4)
rf
Call:
randomForest(formula = Kelas ~ ., data = data.rf1, mtry = 4)
Type of random forest: classification
Number of trees: 500
No. of variables tried at each split: 4

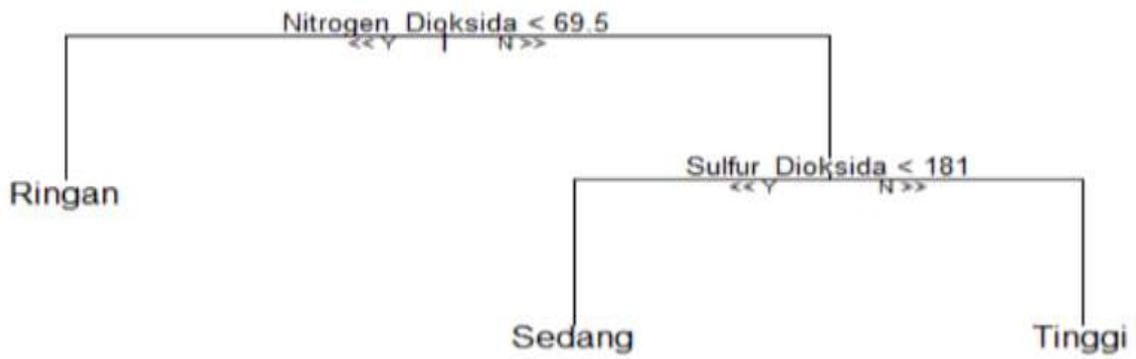
OOB estimate of error rate: 10%
Confusion matrix:
Ringan Sedang Tinggi class.error
Ringan 7 0 0 0.0000000
Sedang 0 10 0 0.0000000
Tinggi 0 1 2 0.3333333
getTree(rf, labelVar = T)
left daughter right daughter split var split point
status prediction
1 2 3 Partikulat 45.5
1 <NA>
2 0 0 <NA> 0.0 -
1 Ringan
3 4 5 Sulfur_Dioksida 183.5
1 <NA>
4 0 0 <NA> 0.0 -
1 Sedang
5 0 0 <NA> 0.0 -
1 Tinggi
rf$predicted
1 2 3 4 5 6 7 8 9
10 11 12
Sedang Ringan Sedang Ringan Sedang Ringan Sedang Ringan Sedang
Ringan Sedang Sedang
13 14 15 16 17 18 19 20
Ringan Sedang SedangSedangRingan Tinggi Sedang Tinggi
Levels: Ringan Sedang Tinggi

rf$importance
MeanDecreaseGini
Sulfur_Dioksida 6.6188727
Nitrogen_Dioksida 3.7752430
Opasitas 0.1101262
Partikulat 1.0301581

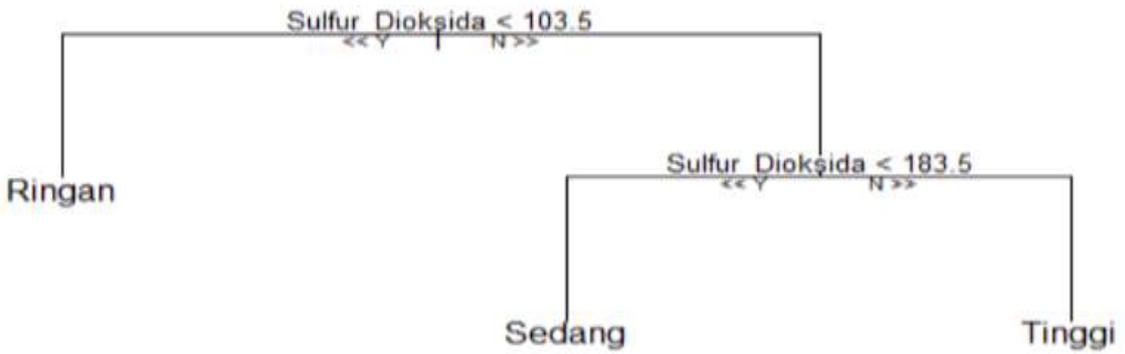
```

Dalam pemodelan *random forest* ditentukan bahwa setiap percobaan pemilihan split keempat variabel dibandingkan (nilai mtry = 4) dan dibentuk 500 pohon. Langkah awal dalam pemodelan adalah dengan melakukan resampel data (*bostraping*) dengan ukuran sampel 20. Selanjutnya berdasarkan data tersebut disusun pohon keputusan dengan metode CART.

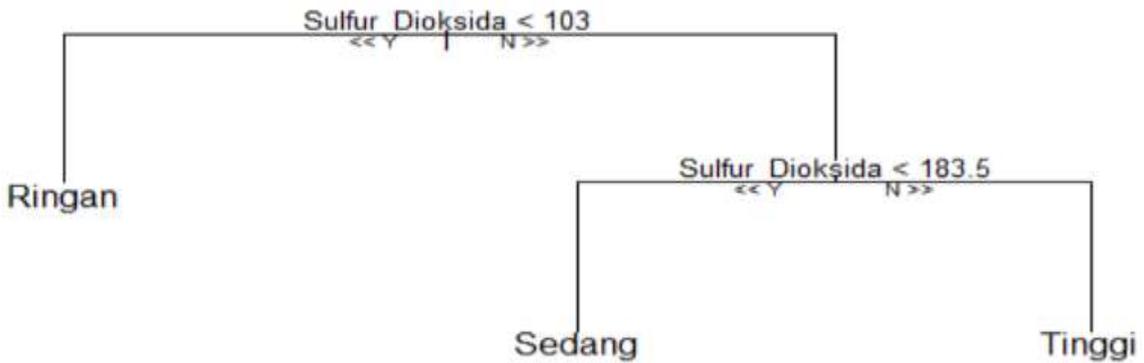
Dari 500 pohon yang di bentuk berikut ini adalah 5 plot pohon keputusan yang didapat :



Gambar 2 Plot Pohon Keputusan 1



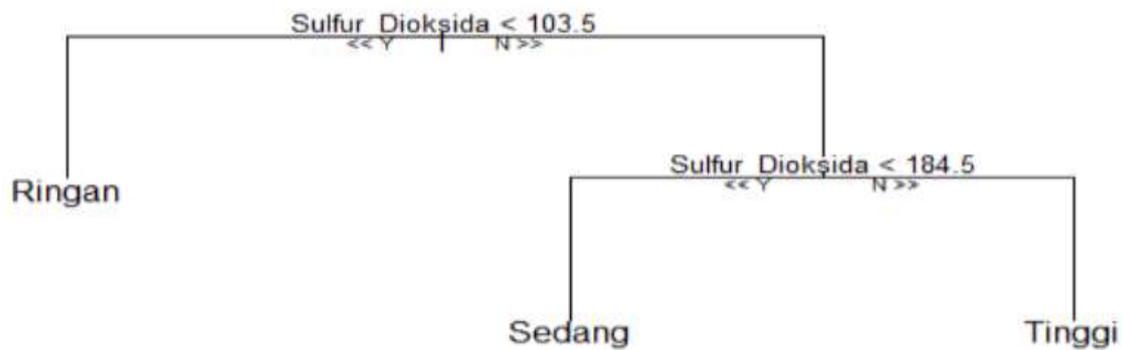
Gambar 3 Plot Pohon Keputusan 2



Gambar 4 Plot Pohon Keputusan 3



Gambar 5 Plot Pohon Keputusan 4



Gambar 6 Plot Pohon Keputusan 5

Berdasarkan hasil prediksi dari model *random forest* menggunakan *software* R Studio dapat disajikan dalam tabel berikut :

Tabel 3 Perbandingan Hasil Aktual dan Prediksi Random Forest

No.	Aktual	Prediksi
1	Sedang	Sedang
2	Ringan	Ringan
3	Sedang	Sedang
4	Ringan	Ringan
5	Sedang	Sedang
6	Ringan	Ringan
7	Sedang	Sedang
8	Ringan	Ringan
9	Tinggi	Sedang
10	Ringan	Ringan
11	Sedang	Sedang
12	Sedang	Sedang
13	Ringan	Ringan
14	Sedang	Sedang
15	Sedang	Sedang
16	Sedang	Sedang
17	Ringan	Ringan
18	Tinggi	Tinggi
19	Sedang	Sedang
20	Tinggi	Tinggi

Dari tabel 2 di atas diketahui bahwa hasil aktual dan prediksi dari *random forest* terdapat perbedaan pada label tingkat pencemaran. Pada hasil aktual pada industri ke-9 dinyatakan bahwa hasil tingkat pencemaran udara berada pada kategori tinggi sedangkan pada hasil prediksi *random forest* berkategori sedang.

#### A. Tingkat Akurasi

Berdasarkan pemodelan tersebut diperoleh hasil matriks konfusi dalam tabel berikut :

Tabel 4 Matriks Konfusi

Aktual	Prediksi		
	Ringan	Sedang	Tinggi
Ringan	7	0	0
Sedang	0	10	0
Tinggi	0	1	2

Berdasarkan tabel 4 diatas dapat dihitung nilai akurasi modelnya sebagai berikut :

$$Akurasi = \frac{7 + 10 + 2}{20} = \frac{19}{20} = 0,95$$

Nilai akurasi yang diperoleh adalah 0,95 atau 95% prediksi model benar.

#### B. Pengaruh Keempat Variabel Independen

Berdasarkan output importance yang didapat dari pemodelan *random forest* dapat disusun dalam tabel 3 berikut :

Tabel 5 Output Importance

Variabel	Nilai Importance
Sulfur dioksida	6,6188
Nitrogen dioksida	3,7752
Opasitas	0,1101
Partikulat	1,0301

Berdasarkan tabel 5 di atas dapat diketahui bahwa variabel yang memiliki 2 pengaruh tertinggi dalam klasifikasi tingkat pencemaran udara pada sektor industri di kota Medan adalah sulfur dioksida (SO<sub>2</sub>) dan Nitrogen Dioksida (NO<sub>2</sub>). Sedangkan variabel 2 terendah adalah opasitas dan partikulat.

## 4. KESIMPULAN

Berdasarkan hasil analisa klasifikasi tingkat pencemaran udara pada sektor industri di kota Medan dengan metode *random forest* yang telah dilakukan dapat diambil kesimpulan :

1. Tingkat akurasi berdasarkan pemodelan *random forest* sebesar 0,95 atau 95% prediksi model benar .
2. Variabel yang memiliki pengaruh tertinggi dalam klasifikasi tingkat pencemaran udara pada sektor industri di kota Medan adalah sulfur dioksida (SO<sub>2</sub>) dan nitrogen dioksida (NO<sub>2</sub>) dan variabel terendah adalah opasitas dan partikulat.

## 5. SARAN

Penelitian ini hanya sampai tahap menerapkan metode *Random Forest* dan menggunakan data industri di kota Medan. Pada pengembangan penelitian selanjutnya dapat dilakukan implementasi berupa pembuatan sistem dengan menggunakan metode terbaik lainnya dan menggunakan data industri terbaru.

### UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih kepada Bapak Saparuddin Nasution dan Ibu Elfina Sembiring orang tua saya yang telah mengasuh dan mendidik saya serta memberikan dukungan moril maupun materil.

Penulis mengucapkan terimakasih kepada Ibu Dr. Fibri Rakhmawati, S.Si., M.Si dan ibu Dr. Riri Syafitri Lubis, S.Pd, M.Si selaku Pembimbing saya yang senantiasa memberikan banyak arahan dan bimbingan kepada penulis dalam menyelesaikan penelitian ini.

### DAFTAR PUSTAKA

- [1]. Wardani, W., Khadiyanto, P., dan Maryono, M., 2003. Pola Persebaran Kualitas Udara Ambient Kawasan Permukiman Di Sekitar Industri Cilegon Sebagai Acuan Ruang Terbuka Hijau Di Kota Cilegon, *Disertasi*, Universitas Diponegoro, Semarang.
- [2]. Peraturan Pemerintah No. 41 Tahun 1999 Tentang Pengendalian Pencemaran Udara. Jakarta.
- [3]. Ramadhan, A., Susetyo, B., dan Indahwati, I. (2019). Penerapan Metode Klasifikasi *Random Forest* Dalam Mengidentifikasi Faktor Penting Penilaian Mutu Pendidikan. *Jurnal Pendidikan dan Kebudayaan*, 4(2), 169-182.
- [4]. Mandra, A. S. 2015. Analisis Pengendalian Pencaeran Emisi Kendaraan Bermotor Di Kabupaten Makassar. *Jurnal Lingkugan*. 3(2): 11-12.
- [5]. Ramadhani, P. (2015). Aplikasi Deteksi Dini Dalam Sistem Sirkulasi Udara Terhadap Polusi Asap, *Disertasi*, Politeknik Negeri Sriwijaya, Palembang.