

Perbandingan Klasifikasi Random Forest, Support Vector Machines, dan LGBM Pada Klasifikasi Kualitas Udara di Jakarta

Comparison of Random Forest, Support Vector Machines, and LGBM Classification for Air Quality Classification in Jakarta

Anisa Ma'u Luthfi*¹, Fatkhurokhman Fauzi²

^{1,2}Statistika, Fakultas Sains dan Teknologi Pertanian, Universitas Muhammadiyah Semarang

Email: ¹anisaluthfi77@gmail.com, ²fatkhurokhmanf@unimus.ac.id

¹Penulis Koresponden

Received: 05 Juni 2024

Accepted: 19 Juli 2024

Published: 01 Agustus 2024



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).
Copyright (c) 2024 JUSTINDO

ABSTRAK

Udara bersih merupakan kebutuhan semua makhluk di bumi guna menunjang keberlangsungan hidup. Kualitas udara di Jakarta merupakan isu serius yang perlu mendapat perhatian serius dari pemerintah dan publik. Ada beberapa faktor utama yang menyebabkan polusi udara di Jakarta. Pertama, industri yang semakin berkembang dengan cepat di wilayah metropolitan ini berkontribusi signifikan terhadap emisi gas buang yang mencemari udara. Selain itu, penambahan jumlah kendaraan bermotor dan mobilitas tinggi juga menyebabkan peningkatan emisi gas buang yang mencemari udara. Kegiatan pembakaran sampah yang tidak teratur dan hujan asam juga turut memperburuk kualitas udara di Jakarta. Penelitian sebelumnya menunjukkan bahwa klasifikasi kualitas udara berdasarkan data Indeks Standar Pencemaran Udara (ISPU) di DKI Jakarta dengan dataset menunjukkan hasil penelitian dari kedua algoritma yang digunakan yakni metode *Support Vector Machine* memiliki akurasi yang lebih baik dalam melakukan klasifikasi dibandingkan dengan *K-Nearest Neighbor* dengan nilai akurasi pada SVM sebesar 98% sedangkan KNN sebesar 96%. Pada penelitian dilakukan Nugroho dan Kawan-kawan didapatkan hasil penelitian dengan dataset data Indeks Standar Pencemaran Udara (ISPU) di DKI Jakarta dengan algoritma *Random Forest* didapatkan hasil akurasi sebesar 90%. Sehingga sebagai bahan perbandingan serta untuk memilih model terbaik dan meningkatkan akurasi model untuk penelitian-penelitian sebelumnya maka di penelitian ini dilakukan perbandingan algoritma antara SVM, *Random Forest*, dan LGBM. Hasil percobaan menunjukkan bahwa tingkat akurasi model yang menggunakan *random forest* memiliki akurasi tertinggi sebesar 98%.

Kata kunci: klasifikasi, random forest, SVM, LGBM, kualitas udara

ABSTRACT

Clean air is a necessity for all creatures on earth to sustain life. Air quality in Jakarta is a serious issue that needs serious attention from the government and the public. There are several main factors that cause air pollution in Jakarta. First, the fast-growing industries in this metropolitan area contribute significantly to exhaust emissions that pollute the air. In addition, the increase in the number of motorized vehicles and high mobility also causes an increase in exhaust emissions that pollute the air. Irregular waste burning activities and acid rain also contribute to worsening air quality in Jakarta. Previous research shows that air quality classification based on Air Pollution Standard Index (ISPU) data in DKI Jakarta with datasets shows the results of research from the two algorithms used, namely the Support Vector Machine method has better accuracy in classifying compared to K-Nearest Neighbor with an accuracy value on SVM of 98% while KNN is 96%. In research conducted by Nugroho and friends, the results of research with the Air Pollution Standard Index (ISPU) data set in DKI Jakarta with the Random Forest algorithm obtained an accuracy of 90%. So as a comparison material and to choose the best model and improve the accuracy of the model for previous studies, in this

study a comparison of algorithms between SVM, Random Forest, and LGBM was carried out. The experimental results showed that the accuracy level of the model using random forest had the highest accuracy of 98%.

Keywords: *clasification, random forest, SVM, LGBM, air quality*

1. Pendahuluan

Salah satu faktor yang penting dalam kebutuhan sehari-hari setiap makhluk yang ada di bumi adalah udara. Jika memiliki kualitas udara yang baik maka diperoleh juga manfaat yang baik bagi makhluk hidup. Kualitas udara di Jakarta merupakan isu serius yang perlu mendapat perhatian serius dari pemerintah dan publik. Tingginya tingkat polusi udara di ibu kota telah memberikan dampak negatif terhadap kesehatan masyarakat serta lingkungan di sekitarnya. Menurut survei yang dilakukan oleh Organisasi Kesehatan Dunia (WHO), Jakarta adalah salah satu kota dengan tingkat polusi udara tertinggi di dunia, yang sering kali melebihi standar yang ditetapkan oleh Badan Perlindungan Lingkungan Amerika Serikat (EPA) dan WHO itu sendiri. (Aisyah Yuri Oktaviana, 2023)

Dalam beberapa dekade terakhir, kualitas udara Jakarta telah menjadi salah satu masalah lingkungan paling penting. Beberapa penyebab penurunan kualitas udara di kota ini termasuk peningkatan jumlah kendaraan bermotor, tingkat konstruksi yang tinggi, dan industrialisasi yang terus meningkat. Untuk membuat kebijakan yang efektif untuk mengurangi dampak negatif polusi udara terhadap kesehatan masyarakat dan lingkungan, pengukuran dan klasifikasi kualitas udara yang tepat sangat penting.

Penurunan kualitas udara secara terus menerus menyebabkan berkurangnya kebutuhan udara yang layak untuk dihirup. Elemen-elemen dalam dataset seperti pm10, pm25, no2, so2, o3, no2 mengenai kualitas udara di DKI Jakarta dapat digunakan untuk menentukan apakah kualitas udara tersebut memenuhi persyaratan Kesehatan atau tidak. Dataset ini didapatkan dari Open Data DKI Jakarta yang berisi karakteristik senyawa kandungan udara yang merupakan faktor standar kualitas udara bersih yang berdampak pada penurunan kualitas udara Jakarta dan dapat digunakan sebagai data untuk klasifikasi kualitas udara.

Proses dalam klasifikasi kualitas udara suatu daerah dapat dilakukan dengan berbagai metode klasifikasi. Penelitian sebelumnya menunjukkan bahwa klasifikasi kualitas udara berdasarkan data Indeks Standar Pencemaran Udara (ISPU) di DKI Jakarta dengan dataset menunjukkan hasil penelitian dari kedua algoritma yang digunakan yakni metode *Support Vector Machine* memiliki akurasi yang lebih baik dalam melakukan klasifikasi dibandingkan dengan *K-Nearest Neighbor* dengan nilai akurasi pada SVM sebesar 98% sedangkan KNN sebesar 96% (Jayadi, Handhayani and Lauro, 2023). Pada penelitian dilakukan Nugroho dan Kawan-kawan didapatkan hasil penelitian dengan dataset data Indeks Standar Pencemaran Udara (ISPU) di DKI Jakarta dengan algoritma *Random Forest* didapatkan hasil akurasi sebesar 90% (Nugroho, Asror and Wibowo, 2022).

2. Metode Penelitian



Gambar 1. Urutan Penelitian

Metode penelitian ini dilaksanakan untuk mengetahui kualitas udara di Jakarta. Data yang digunakan merupakan data dari Dinas Lingkungan Hidup DKI Jakarta yang tersedia di Open Data.

- Data yang sudah didapatkan akan dilakukan proses *data cleaning*. Proses pembersihan data dari *missing value* atau data yang kosong serta dari data yang *duplikat* disebut *Data Cleaning*.
- Setelah dilakukannya *Data Cleaning* Langkah selanjutnya dilakukan *Transformasi Data* yakni teknik perubahan kategori data yang awalnya data berbentuk kategori menjadi data yang berbentuk numerik.

- Hasil data yang telah didapatkan dalam *Data Cleaning* selanjutnya akan digunakan sebagai input data pada klasifikasi kualitas udara menggunakan *Random Forest*, *SVM*, dan juga *LGBM*.

2.1. Data Penelitian

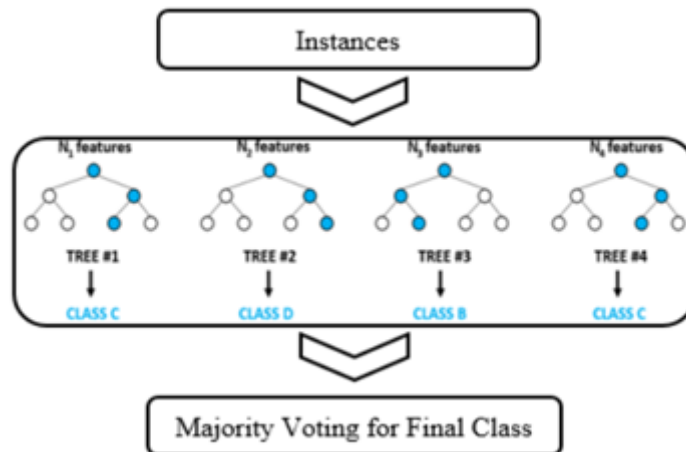
Dataset yang digunakan dalam penelitian ini merupakan dataset sekunder yang didapatkan dari Open Data DKI Jakarta [Home || Satudata Jakarta](#) yang terdiri dari 21.915 baris serta 11 kolom. Pada penelitian ini dataset yang digunakan tentang kualitas udara di DKI Jakarta yang terdiri dari beberapa variabel, seperti tanggal, stasiun, pm10, pm25, co, so2, o3, no2, max, critical, dan kategori.

Tabel 1. Penjelasan Variabel

Column Name	Description
Date	Tanggal pengukuran Kualitas Udara
Stations	Lokasi pengukuran di stasiun
PM10	Materi partikulat mikrometer
SO2	Sulfur dioksida
CO	Karbon monoksida
O3	Ozon
NO2	Nitrogen dioksida
Max	Nilai terukur tertinggi
Critical	Parameter terukur tertinggi
Category	Kategori perhitungan ISPU
PM2.5	Partikulat 2,5 mikrometer

2.2. Random Forest

Random Forest merupakan teknik model prediktif yang mencakup teknik modeling yang bekerja sama. Dalam Bahasa gampangnya Random forest merupakan gabungan dari beberapa decision tree dimana dari semua decision tree yang didapatkan dalam model saling bekerja sama untuk saling meningkatkan kinerja model. Di Dalam random forest terdapat ide model ansambel dimana suatu model bekerja sama untuk memecahkan suatu masalah. Dengan itu maka Random Forest memiliki Tingkat keberhasilan yang lebih tinggi dibandingkan model lainnya. (Praneswara, 2023)

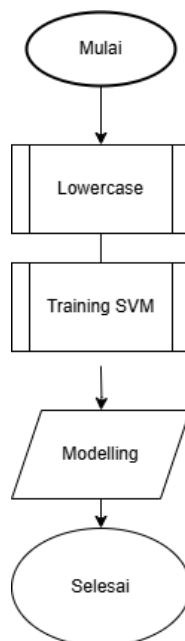


Gambar 2. Ilustrasi Random Forest

Parameter yang digunakan dalam Random Forest terdiri dari beberapa parameter salah satunya seperti $n_estimator$ yang merupakan representasi dari banyaknya jumlah tree yang digunakan, biasanya makin banyak tree yang digunakan makin bagus hasil akurasi yang didapatkan. (Widjiyati, 2021)

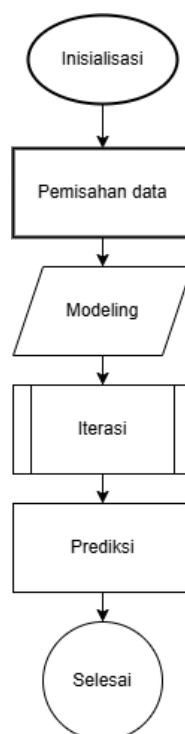
2.3. Support Vector Machine (SVM)

Support Vector merupakan data yang paling dekat dengan hyperlane. Sedangkan SVM sendiri merupakan suatu metode klasifikasi yang menggunakan prinsip mencari hyperlane. Garis yang memisahkan data sendiri juga merupakan hyperlane. Sehingga secara keseluruhan secara garis besar adalah metode klasifikasi yang digunakan untuk mendapatkan hasil akhir dari pembuatan system yang memiliki prinsip untuk mencari hyperlane yang memiliki margin terbesar. (Irmada and Astriratma, 2020)



Gambar 3. Ilustrasi SVM

2.4. LGBM



Gambar 4. Ilustrasi LGBM

Sumber LGBM (Light Gradient Boosting Machine) adalah kerangka Peningkatan gradien yang menggunakan algoritma pembelajaran berdasarkan pohon keputusan. Bisa dikatakan penggunaan memori yang rendah, akurasi yang lebih tinggi, dukungan pembelajaran paralel, dan dapat menangani data skala besar. Dibandingkan dengan algoritma pembelajaran mesin pada umumnya, kecepatannya sangat cepat. LGBM menggunakan algoritma histogram. Ide dasar dari algoritma histogram adalah untuk mendiskritisasi nilai eigen floating[1]point kontinu menjadi k bilangan bulat, dan pada saat yang sama membuat histogram dengan lebar k. Saat melintasi data, gunakan nilai yang di diskritisasi sebagai indeks untuk mengumpulkan statistik dalam histogram. Setelah melintasi data satu kali, histogram mengumpulkan statistik yang diperlukan, dan kemudian melintasi untuk menemukan nilai optimal sesuai dengan nilai diskrit histogram.(Seshagiri Rao and Priyanka, no date).

3. Hasil Dan Pembahasan

3.1. Preprocessing Data

Dalam penelitian ini menggunakan dataset ISPU DKI Jakarta yang terdiri dari 21.915 baris dan 11 variabel atau kolom, Seperti yang ditunjukkan pada Gambar 5. Langkah awal dari penelitian ini adalah preprocessing data yang dilakukan dalam penelitian ini preprocessing data melalui tahap Data Cleaning dan juga Transformasi Data.

Tahapan data cleaning merupakan tahapan dilakukannya pengecekan dan penanganan missing value.Data cleaning yang dilakukan dalam dataset ini yakni penghapusan fitur pm25 dikarenakan lebih dari 80% data merupakan missing value.Dan untuk penanganan missing value pada fitur lainnya yakni dengan mengisi data missing dengan modus untuk tipe data kategorik dan median untuk tipe data numerik.

tanggal	stasiun	pm10	so2	co	o3	no2	max	critical	kategori	pm25
01/01/2010	DKI 1	60	4	73	27	14	73	CO	SEDANG	Nan
02/01/2010	DKI 1	32	2	16	33	9	33	O3	BAIK	Nan
03/01/2010	DKI 1	27	2	19	20	9	27	PM10	BAIK	Nan
04/01/2010	DKI 1	22	2	16	15	6	22	PM10	BAIK	Nan
05/01/2010	DKI 1	25	2	17	15	8	25	PM10	BAIK	Nan
...
27/12/2021	DKI 5	54	36	14	21	47	76	PM25	SEDANG	76
28/12/2021	DKI 5	44	20	11	21	33	68	PM25	SEDANG	68
29/12/2021	DKI 5	34	28	8	25	29	54	PM25	SEDANG	54
30/12/2021	DKI 5	53	25	15	23	44	75	PM25	SEDANG	75
31/12/2021	DKI 5	60	28	19	30	53	87	PM25	SEDANG	87

Gambar 5.Dataset ISPU

Tahapan selanjutnya yakni Transformasi data, data cleaning yang sudah dilakukan penanganan missing value langkah selanjutnya yakni Transformasi data yang dilakukan pada fitur kategori.Transformasi data dilakukan agar dapat dianalisis oleh model klasifikasi. Fitur yang dibutuhkan dalam Transformasi data adalah kategori yang akan diubah dengan menggunakan label encoder, yakni mengubah data kategorik pada kolom kategori menjadi numerik dengan perubahan nilai yaitu BAIK diubah menjadi 0, SEDANG diubah menjadi 1, TIDAK SEHAT diubah menjadi 2, SANGAT TIDAK SEHAT diubah menjadi 3, BERBAHAYA diubah menjadi 4.

3.2 Pemrosesan Data

Pemrosesan data dilakukan dengantools pemogramann phyton dengan digunakan beberapa library seperti:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
import lightgbm as lgb
from sklearn.svm import SVC
```

Gambar 6. Library Python

Pemrosesan data ini terdiri dari beberapa tahapan seperti

- Proses pembersihan data dari *missing value* atau data yang kosong serta dari data yang *duplikat* disebut *Data Cleaning*.

```
# Mengisi nilai numerik dengan median
df_clean[['pm10', 'so2', 'co', 'o3', 'no2', 'max']].fillna(df_clean[['pm10', 'so2', 'co', 'o3', 'no2', 'max']].median())

# Mengisi nilai kategorik modus
df_clean['critical'] = df_clean['critical'].fillna(df_clean['critical'].mode().iloc[0])
```

Gambar 7. Proses Data Cleaning

- *Transformasi Data* yakni teknik perubahan kategori data yang awalnya data berbentuk kategori menjadi data yang berbentuk numerik.

```
#Ubah data kategorik kategori menjadi numerik
df_clean["kategori"] = df_clean["kategori"].map({"BAIK":0, "SEDANG":1, "TIDAK SEHAT":2, "SANGAT TIDAK SEHAT":3, "BERBAHAYA":4})
df_clean
```

	tanggal	stasiun	pm10	so2	co	o3	no2	max	critical	kategori
0	2010-01-01	DKI1 (Bunderan HI)	60.0	4.0	73.0	27.0	14.0	73.0	CO	1
1	2010-01-02	DKI1 (Bunderan HI)	32.0	2.0	16.0	33.0	9.0	33.0	O3	0
2	2010-01-03	DKI1 (Bunderan HI)	27.0	2.0	19.0	20.0	9.0	27.0	PM10	0
3	2010-01-04	DKI1 (Bunderan HI)	22.0	2.0	16.0	15.0	6.0	22.0	PM10	0
4	2010-01-05	DKI1 (Bunderan HI)	25.0	2.0	17.0	15.0	8.0	25.0	PM10	0
...
21910	2021-12-27	DKI5 (Kebon Jeruk)	54.0	36.0	14.0	21.0	47.0	76.0	PM25	1
21911	2021-12-28	DKI5 (Kebon Jeruk)	44.0	20.0	11.0	21.0	33.0	68.0	PM25	1
21912	2021-12-29	DKI5 (Kebon Jeruk)	34.0	28.0	8.0	25.0	29.0	54.0	PM25	1
21913	2021-12-30	DKI5 (Kebon Jeruk)	53.0	25.0	15.0	23.0	44.0	75.0	PM25	1
21914	2021-12-31	DKI5 (Kebon Jeruk)	60.0	28.0	19.0	30.0	53.0	87.0	PM25	1

Gambar 8. Transformasi Data

Data yang telah selesai melalui preprocessing data selanjutnya akan dilakukan pemodelan. Tahapan pemodelan akan dilakukan melalui tiga metode yaitu Random Forest, SVM, dan LGBM.

Sebelum dilakukannya klasifikasi data yang akan digunakan harus sudah melalui preprocessing data. Data akan dibagi menjadi data training dan data testing dengan perbandingan 80% atau sebanyak 13.692 dan 20% atau sebanyak 3.423. Data training akan diuji menggunakan klasifikasi machine learning dengan model Random Forest, SVM, dan LGBM. Sedangkan data testing akan digunakan untuk data pengujian performa ketiga metode yang akan digunakan.

Metode dibangun dengan fitur independent pm10, so2, co, o3, no2, dan dengan fitur dependent kategori. Maka dengan percobaan metode tersebut sehingga dapat diperoleh model terbaik

3.4 Pengujian

Penelitian ini menggunakan skema penelitian split data. Dimana dataset yang telah dilakukan preprocessing data akan di split dengan banyaknya data train sebanyak 13.692 dan banyaknya data

testing sebanyak 3.423. Pengujian akan dilakukan dengan tiga model klasifikasi yaitu Random Forest, SVM, dan LGBM.

Hasil dari analisis berupa *confusion matrix*, *True Positive*(TP), *True Negative*(TN), *False Positif* (FP), dan *False Negative* (FN). Dan berdasarkan hal tersebut didapatkan juga data yang berguna untuk melihat peforma model yang berupa hasil akurasi, presisi dan recall. (Adityo, no date)

- *Confusion matrix* adalah table matrix yang berfungsi sebagai penunjuk peforma algoritma.
- *True Positive* adalah banyaknya data, Dimana data yang aslinya positif dan di prediksi dengan positif.
- *True Negative* adalah banyaknya data yang nilai data aslinya negative dan di prediksi negative.
- *False Negative* adalah banyaknya data yang nilai data aslinya positif diprediksi dengan nilai negative.
- *False Positif* adalah banyaknya data yang nilai data aslinya negative tetapu diprediksi positif oleh model.
- Akurasi adalah rasio prediksi yang benar, baik positif maupun berdampak negatif pada semua data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Presisi adalah rasio prediksi positif yang benar terhadap semua prediksi positif.

$$Precision = \frac{TP}{TP + FP}$$

- Recall adalah rasio prediksi positif yang benar terhadap semua data adalah positif.

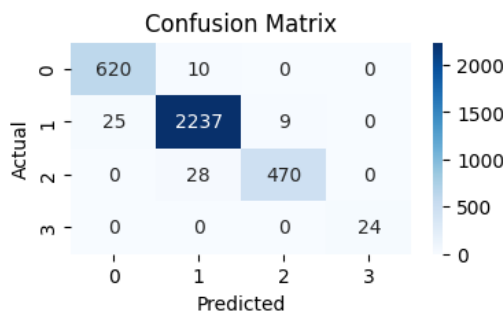
$$Recall = \frac{TP}{TP + FN}$$

Pengujian menggunakan metode *Random Forest* dengan banyaknya *n_estimator*=100 didapatkan hasil accuracy, presisi, dan recall seperti yang ditunjukkan dalam Tabel 1.

Tabel 1. Hasil pengujian menggunakan Random Forest

Accuracy	Presisi	Recall
98%	96%	98%

Pada Tabel 1 dapat dilihat hasil dari pengujian dengan metode *Random Forest* didapatkan hasil akurasi sebesar 98%, presisi sebesar 96%, dan recall sebesar 98%. Dan dari model *random forest* didapatkan table *confusion matrix* sebagai berikut:



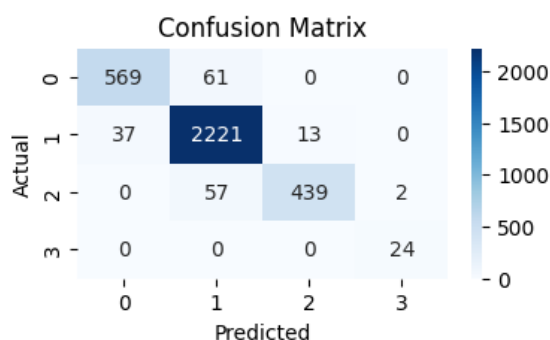
Gambar 9. Confusion matrix random forest

Pengujian menggunakan metode SVM dan didapatkan hasil accuracy, presisi, dan recall seperti yang ditunjukkan dalam Tabel 2.

Tabel 2. Hasil pengujian menggunakan SVM

Accuracy	Presisi	Recall
95%	94%	90%

Pada Tabel 2 dapat dilihat hasil dari pengujian dengan metode SVM didapatkan hasil akurasi sebesar 95% hal ini berbeda dari penelitian sebelumnya dikarenakan adanya perbedaan tahun yang digunakan, pada penelitian sebelumnya menggunakan data tahun 2021 dan pada penelitian ini menggunakan data tahun 2010 hingga 2021, presisi sebesar 94%, dan recall sebesar 98%. Dan dari model SVM didapatkan tabel *confusion matrix* sebagai berikut:



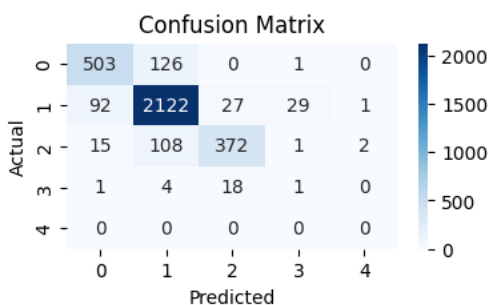
Gambar 10. Confusion matrix SVM

Pengujian menggunakan metode LGBM dan didapatkan hasil accuracy, presisi, dan recall seperti yang ditunjukkan dalam Tabel 3.

Tabel 3. Hasil pengujian menggunakan SVM

Accuracy	Presisi	Recall
88%	82%	80%

Pada Tabel 2 dapat dilihat hasil dari pengujian dengan metode LGBM didapatkan hasil akurasi sebesar 88%, presisi sebesar 82%, dan recall sebesar 80%. Dan dari model LGBM didapatkan tabel *confusion matrix* sebagai berikut:



Gambar 11. Confusion matrix LGBM

4. Evaluasi Dan Analisis

Perbandingan performa hasil antara tiga metode yang digunakan antara Random Forest, SVM, dan LGBM. Disajikan pada Tabel 4 ketiga metode memiliki akurasi yang berbeda tetapi tidak jauh berbeda.

Tabel 4. Perbandingan Hasil

Metode	Akurasi	Presisi	Recall
Random Forest	98%	96%	98%
SVM	95%	94%	90%
LGBM	88%	82%	80%

Hasil perbandingan secara keseluruhan dapat disimpulkan yaitu hasil secara keseluruhan menunjukkan bahwa performa metode Random Forest lebih baik dari metode SVM dan LGBM. Berdasarkan hasil tersebut maka dapat disimpulkan bahwa metode klasifikasi terbaik secara keseluruhan adalah Random Forest dengan akurasi sebesar 98% dan estimator 100.

5. Kesimpulan

Udara bersih merupakan kebutuhan primer bagi kehidupan makhluk di bumi. Semakin baik kualitas udara maka semakin baik juga kehidupan makhluknya. Kualitas udara di Jakarta merupakan isu serius yang perlu mendapat perhatian serius dari pemerintah dan publik. Tingginya tingkat polusi udara di ibu kota telah memberikan dampak negatif terhadap kesehatan masyarakat serta lingkungan di sekitarnya. Penurunan kualitas udara menyebabkan urgensi kebutuhan dalam mengawasi kebutuhan udara yang layak untuk dihirup. Kualitas udara yang sesuai dengan standar kesehatan dapat diketahui melalui zat-zat yang terkandung didalamnya. Proses klasifikasi kualitas udara dengan metode Random Forest, SVM, dan LGBM. Hasil penelitian secara keseluruhan yaitu menunjukkan tingkat performa dari metode Random Forest memiliki akurasi tertinggi sebesar 98% yang merupakan model klasifikasi terbaik dengan presisi sebesar 96% artinya model mampu memprediksi 96% dari total prediksi positif yang benar, dan nilai recall sebesar 98% artinya dari seluruh data yang positif model Random Forest mampu memprediksi sebesar 98%. Sedangkan metode SVM memiliki akurasi 95%, dan metode LGBM memiliki akurasi sebesar 88%.

Daftar Pustaka

- Adityo, N.A.F.A.W. (no date) *Klasifikasi Tingkat Kualitas Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma Random Forest*.
- Aisyah Yuri Oktavania (2023) *Data Polusi Udara di Indonesia 2015-2023, Penyebab, & Dampaknya*. Available at: <https://tirto.id/info-data-polusi-udara-di-indonesia-pada-2015-2023-penyebabnya-gPhD> (Accessed: 2 July 2024).
- Azimah, F. and Wardani, K.R.N. (2022) "Klasifikasi Deteksi Gejala Awal COVID-19 Dengan Metode Logistic Regression, Random Forest Classifier dan Support Vector Machine," *Jurnal Locus Pengabdian dan Penelitian* [Preprint]. Available at: <https://doi.org/10.58344/locus.v1i6.135> (Accessed: June 4, 2024).
- Efendi, A. et al. (2022) "Klasifikasi Kebakaran Hutan Riau Menggunakan Random Forest dan Visualisasi Citra Sentinel-2." Available at: <https://doi.org/10.30865/klik.v4i3.1521> (Accessed: June 4, 2024).

- Filemon, B., Mawardi, V.C. and Perdana, N.J. (2022) "PENGUNAAN METODE SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI SENTIMEN E-WALLET," *Jurnal Ilmu Komputer dan Sistem Informasi* [Preprint].
- Irmada, H.N. and Astriratma, R. (2020) "Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM)," *Jurnal Resti* [Preprint]. Available at: <https://doi.org/10.29207/resti.v4i5.2313> (Accessed: June 4, 2024).
- Jayadi, B.V., Handhayani, T. and Lauro, M.D. (2023) "PERBANDINGAN KNN DAN SVM UNTUK KLASIFIKASI KUALITAS UDARA DI JAKARTA," *Jurnal Ilmu Komputer dan Sistem Informasi*, Vol. 11 No. 2, pp. 1–7. Available at: <https://doi.org/10.24912/jiksi.v11i2.26006> (Accessed: June 4, 2024).
- Nugroho, A., Asror, I. and Wibowo, Y.F.A. (2022) . "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma Random. ," *eProceedings of Engineering* [Preprint], (<https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/issue/view/210>). Available at: - (Accessed: June 4, 2024).
- Praneswara, A.O. (2023) . "Perbandingan K-Nearest Neighbors, Support Vector Dan Random Forest Pada Prediksi Medical Cost.," *Indonesian Journal of Computer Science* [Preprint]. Available at: <https://doi.org/10.33022/ijcs.v12i4.3298> (Accessed: June 4, 2024).
- Putra, R.E., Kalista, M. and Setianingsih, C. (2023) "Klasifikasi prediksi kualitas udara Menggunakan metode Support Vector Machine (SVM)," *eProceedings of Engineering* [Preprint], (<https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/issue/view/218>) (Accessed: June 4, 2024).
- Seshagiri Rao, M.K. and Priyanka, M.K. (no date) *DIABETES PREDICTION USING MACHINE LEARNING ALGORITHMS LIKE SVM, NB AND LGBM*. Available at: www.anveshanaindia.com.
- Wahyudiyanta, S.A. and Supriyati (2024) "Analisis Kualitas Udara Jakarta dan Prediksi Tingkat Polusi dengan Metode Mesin Pembelajaran SVM," *Seminar Nasional Sains dan Teknologi (SAINTEK)* [Preprint], (<https://jurnal.pelitabangsa.ac.id/index.php/SAINTEK/issue/view/215>).
- Widjiyati, N. (2021) " Implementasi Algoritme Random Forest Pada Klasifikasi Dataset Credit Approval," *Jurnal Janitra Informatika dan Sistem Informasi*, 1. Available at: <https://doi.org/10.25008/janitra.v1i1.118> (Accessed: June 4, 2024).
- Yuli Endra, R. and Saputra, G.A. (2022) "IMPLEMENTASI FUZZY INFERENCE SYSTEM (FIS) METODE TSUKAMOTO UNTUK MONITORING KUALITAS UDARA," *Jurnal Komputasi* [Preprint]. Available at: <https://doi.org/10.23960/komputasi.v10i1.2962> (Accessed: June 4, 2024).