

PREDIKSI PARTICULAT MATTER 10 MENGGUNAKAN MODEL DECISION TREE REGRESSOR DI DKI JAKARTA

Muhammad Zaidan Hakim ^{a,1}, Safri Emanuel Manullang ^{a,2}, Agung Hari Saputra ^{a,3}

^aSekolah Tinggi Meteorologi Klimatologi dan Geofisika, Indonesia

Correspondent Author: Safri Emanuel Manullang (email: safriemanuel@gmail.com)

ABSTRACT

Jakarta is one of the cities with high levels of pollution. This is a special concern for the Indonesian Government, good air quality will have a positive impact on life. Air pollution modeling prediction is carried out by using past air pollution data, current conditions, and scientific principles to estimate future pollution conditions. This study was conducted to predict PM10 parameters in the DKI Jakarta area using the Decision Tree Regressor (DTR) model. Decision Tree Regressor is a simple representation of a classification technique for a number of finite class models (sets that have a number of members that can be counted explicitly), to test the error of the results obtained from the implementation of Decision Tree Regression using MAE, RMSE, and MAPE from the results of the DTR model implementation. The results of the PM10 modeling obtained an RMSE value of 7.5970, MAE of 5.4514, and MAPE of 0.2857. From the results obtained, it can be concluded that the Decision Tree Regressor model is quite good at predicting PM10 in DKI Jakarta.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Article History

Received 2024-01-10

Revised 2024-01-27

Accepted 2024-06-08

Keywords

Prediksi

Particulate Matter (PM10)

DKI Jakarta

Decision Tree Regressor

Pendahuluan

DKI Jakarta adalah ibu kota dan pusat pemerintahan Republik Indonesia, sebagai pusat politik, ekonomi, dan budaya negara, DKI Jakarta memiliki peran yang sangat signifikan. Dengan perannya yang sangat sentral, DKI Jakarta menjadi representasi keberagaman, kompleksitas, dan dinamika Indonesia sebagai sebuah negara besar di Asia Tenggara[1]. Jakarta dikenal juga dengan kota metropolitan yang memiliki kandungan polusi udara yang tinggi. Udara adalah salah satu kebutuhan utama bagi semua makhluk hidup di muka bumi. Kualitas udara yang baik akan berdampak positif bagi kehidupan begitupun sebaliknya. Manusia harus pintar dalam menjaga kualitas udara dengan mengenal dampak dan pengaruh udara di berbagai sektor.

Pencemaran udara merupakan salah satu permasalahan kesehatan masyarakat yang mempunyai dampak serius pada kesehatan manusia dan kualitas lingkungan. Umumnya terjadi di negara berkembang dan mempunyai banyak industri, terutama dihasilkan oleh lalu lintas kendaraan, aktivitas industri dan pemanas rumah selama musim dingin [2]. Banyak penelitian yang dilakukan di belahan dunia membuktikan bahwa pencemaran udara mengakibatkan gangguan kesehatan pada manusia. Dampak Kesehatan yang dihasilkan yang disebabkan oleh tingginya sebaran pm10 adalah kematian dini yang berhubungan dengan penyakit yang dapat disebabkan oleh sebaran pm10 yang tinggi. Penyakit yang diidap berupa penyakit asma, ISPA, penyakit saluran pernafasan dan bronkitis kronis [3]. Pencemaran udara dapat terjadi di luar ruangan (*outdoor*) maupun di dalam ruangan (*indoor*). *Particulate Matter* (PM10) memiliki tingkat toksisitas yang tinggi dan mempunyai peran dalam rusaknya udara

ambient, debu mengandung partikel zat padat dan cair yang dapat ditemukan di udara dan menimbulkan gangguan saluran pernafasan dan mencemari udara [3].

Penyebab pencemaran udara terbagi menjadi dua adalah polutan berbentuk partikel dan gas-gas. Partikel pencemar dapat berupa total *suspended particulate*/partikel tersuspensi total (TSP) dengan ukuran diameter partikel sampai dengan 100 μ m, partikel berdiameter kurang dari 10 μ m (PM10), dan partikel berdiameter kurang dari 2.5 μ m (PM 2,5); sedangkan gas-gas pencemar dapat berupa sulfur dioksida (SO₂), nitrogen dioksida (NO₂), karbon monoksida (CO), oksidan/ozon permukaan (O₃), dan lainnya (Yulinawati, 2015). Partikulat (PM10) adalah Partikel udara yang berukuran lebih kecil dari 10 mikron (mikrometer) [4]. PM10 dapat berasal dari berbagai sumber, termasuk industri, transportasi, pembakaran bahan bakar, pembakaran sampah, dan aktivitas manusia lainnya [4]. Berbagai jenis zat pencemar udara, benda partikulat atau *particulate matter* berdiameter 10 mikron (PM10) mendapatkan perhatian khusus karena dinilai memiliki pengaruh lebih besar terhadap gangguan kesehatan manusia dibandingkan dengan zat-zat pencemar lainnya. WHO (2011), juga menyatakan bahwa PM10 merupakan prediktor kesehatan yang baik. Perkembangan kawasan perkotaan secara fisik dapat menghilangkan hakikat ruang terbuka dan luas, dan secara ekologis dapat menimbulkan berbagai gangguan proses alam di lingkungan perkotaan, yaitu kenaikan suhu, frekuensi banjir, yang dapat berdampak pada penurunan polusi udara dalam keanekaragaman hayati. Menurut Rawung (2015) "*Eco-cities*" merupakan salah satu inisiatif yang dapat dilakukan untuk melestarikan dan memulihkan ruang terbuka hijau di lingkungan perkotaan agar dapat berperan optimal dari sudut pandang lingkungan, sosial, dan ekonomi.

Prediksi merupakan cara untuk memperkirakan suatu hal secara sistematis yang paling mungkin terjadi dimasa yang akan datang berdasarkan informasi masa lalu dan sekarang. Prediksi (*forecasting*) adalah suatu kegiatan yang memperkirakan apa yang terjadi pada masa akan datang [5]. Prediksi pemodelan cuaca dilakukan dengan memproses data sejarah pencemaran udara, kondisi saat ini, dan prinsip ilmiah untuk memperkirakan kondisi pencemaran udara di masa depan [5][6]. Ini melibatkan penggunaan model matematika yang kompleks untuk mensimulasikan perilaku atmosfer, laut, dan permukaan bumi yang mempengaruhi cuaca. Prediksi pemodelan cuaca sangat bermanfaat bagi berbagai macam bidang di Indonesia, seperti pertanian, transportasi, pariwisata, dan mitigasi bencana.

Penelitian ini dilakukan untuk memprediksi parameter PM10 di wilayah DKI Jakarta dengan model *Decision Tree Regressor*. Model ini digunakan karena mudah digunakan oleh yang bukan ahli dalam prediksi, dengan menyediakan opsi *default* yang kuat dan mengotomatisasi banyak keputusan pemodelan.

Metode

Decision Tree Regressor merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun simpul akar ditandai dengan nama atribut. Model DTR merupakan salah satu algoritma *machine learning* yang dapat digunakan untuk memodelkan data time series dengan beberapa hal yang perlu diperhatikan seperti urutan waktu harus disusun dengan benar, memasukkan informasi waktu kedalam fitur-fitur yang bisa di-*input* ke model seperti mingguan, bulanan, atau musiman [10][11]. Manfaat utama dari penggunaan DTR adalah kemampuannya untuk mem-*break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel. sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan [10].

Mempersiapkan regresi pohon keputusan (DTR) berbasis deret waktu melibatkan serangkaian fase khusus untuk memperhitungkan karakteristik data temporal. Langkah pertama adalah mengumpulkan data deret waktu, termasuk informasi waktu untuk setiap observasi. Pemrosesan data kemudian terjadi, termasuk penyesuaian waktu, penanganan nilai yang hilang, dan mengidentifikasi outlier yang mungkin mempengaruhi analisis. Kemudian, untuk menjamin keakuratan hasil model, dataset dibagi menjadi data latih dan data uji dengan mempertimbangkan time series. Kemudian, model DTR dikembangkan dengan mempertimbangkan aturan pemisahan di pohon keputusan yang sesuai dengan urutan

temporal data. Penentuan parameter model juga disesuaikan dengan karakteristik deret waktu, misalnya dengan memperhitungkan efek penundaan atau penyesuaian kedalaman pohon. Saat melatih model deret waktu, difokuskan pada struktur temporal data dan melakukan evaluasi menggunakan data pengujian yang mencerminkan kondisi masa depan.

Opsi pengoptimalan model dapat dicapai dengan menyesuaikan parameter atau menerapkan teknik pembersihan untuk menghindari penyesuaian berlebihan pada tren atau pola tertentu dalam data. Setelah suatu model ditentukan kesesuaiannya, model tersebut dapat digunakan untuk membuat prediksi data deret waktu baru atau untuk memprediksi nilai masa depan. Validasi model memberikan keyakinan tambahan pada performa yang konsisten, terutama ketika data deret waktu menunjukkan variasi yang signifikan atau dipengaruhi oleh faktor musiman. Dokumentasi model DTR berbasis deret waktu dan hasil evaluasi harus dilakukan dengan hati-hati, menafsirkan bagaimana variabel deret waktu memengaruhi prediksi dan bagaimana perubahan terjadi seiring waktu. Oleh karena itu, persiapan DTR berbasis deret waktu memerlukan langkah-langkah khusus untuk memastikan bahwa model dapat menangani dinamika data temporal secara efektif. Dalam penyusunan model ini dibutuhkan beberapa tahap yaitu.

Persiapan Data

Data yang digunakan adalah data indeks PM10 wilayah DKI Jakarta. Dataset diunduh dari laman (<https://data.jakarta.go.id/dataset/indeks-standar-pencemaran-udara>) dan diolah menggunakan Microsoft Excel. Selain itu, Untuk memungkinkan *plotting* data, data diproses dengan melakukan modifikasi menambahkan data yang hilang, dan mengubah format data (Ivanoti dkk., 2023).

Tabel 1. Sampel Data yang Sudah Diolah Pada Microsoft Excel

Tanggal	pm10
01/01/2020	77.0
02/01/2020	52.0
03/01/2020	38.0

Pembangunan dan Pengujian Model.

Tabel 2. Setup Pembangunan Model

Description	Value
<i>target</i>	pm10
<i>fh</i>	30
<i>session_id</i>	123
<i>fold</i>	5
<i>verbose</i>	True
<i>remove_harmonics</i>	True
<i>seasonal_period</i>	30
<i>fold_strategy</i>	sliding
<i>max_sp_to_consider</i>	48
<i>harmonic_order_method</i>	<i>harmonic_strength</i>
<i>num_sps_to_use</i>	1
<i>coverage</i>	0.7

Setelah dilakukan pengolahan data, dilakukan tahap berikutnya yaitu pembangunan model dengan pengaturan yang tertera pada tabel 2. Setelah melakukan pembangunan model, dilakukan uji kesalahan terhadap hasil yang sudah didapat dari implementasi *Decision Tree Regression*. Uji kesalahan dilakukan menggunakan MAE, RMSE, dan MAPE dari hasil implementasi model DTR sebelumnya untuk mengetahui tingkat keakuratan prediksi model regresi.

RMSE

Root mean square error (RMSE) merupakan suatu jenis kesalahan yang tergantung pada skala dan tidak dapat dibandingkan secara langsung antar skala yang berbeda. Untuk mengatasi hal ini, perbaikan RMSE melibatkan pembuatan akar kuadrat dari perbedaan antara nilai yang diharapkan dan nilai sebenarnya, membentuk dasar yang bersifat umum untuk kualitas yang diantisipasi dan kualitas sebenarnya [12]. Nilai hasil RMSE akan semakin baik jika nilainya mendekati nol.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^N (y_k - \hat{y}_k)^2} \quad (1)$$

y_k merupakan nilai asli, kemudian \hat{y}_k merupakan nilai hasil prediksi, dan N merupakan jumlah sampel data.

MAE

Mean absolute error (MAE) digunakan untuk menghitung nilai rata-rata antara nilai aktual dalam seri waktu dan nilai yang diprediksi. MAE dikenal karena kesalahannya berskala, di mana evaluasi kesalahan tergantung pada nilai skala dan tidak dapat dibandingkan secara langsung antar skala yang berbeda [12]. Hasil yang didapat dari prediksi serta data asli akan digunakan untuk menunjukkan rata-rata kesalahan mutlak berdasarkan metode yang digunakan

$$MAE = \frac{1}{n} \sum_{i=1}^N |y_k - \hat{y}_k| \quad (2)$$

y_k merupakan nilai asli, kemudian \hat{y}_k merupakan nilai hasil prediksi, dan N merupakan jumlah sampel data.

MAPE

Mean Absolute Percentage Error (MAPE) Evaluasi model terakhir untuk diperhitungkan kesalahan atau akurasi dari 86 Jurnal Ilmiah Teknologi dan Rekayasa Volume 28 No. 1 April 2023 hasil model regresi adalah MAPE [13]. MAPE berguna ketika ukuran atau besar variabel ramalan itu penting dalam menguji ketepatan ramalan. MAPE menunjukkan seberapa besar kesalahan dalam meramal yang dibandingkan dengan nilai nyata dalam deret. Pengukuran akurasi MAPE memiliki persamaan sebagai berikut

$$MAPE = \frac{1}{N} \sum_{k=1}^N \left| \frac{y_k - \hat{y}_k}{y_k} \right| \times 100 \quad (3)$$

y_k merupakan nilai asli, kemudian \hat{y}_k merupakan nilai hasil prediksi, dan N merupakan jumlah sampel data. Untuk kategori nilai MAPE dapat dilihat pada tabel 3

Tabel 3. Kriteria nilai MAPE (*Mean Absolute Percentage Error*)

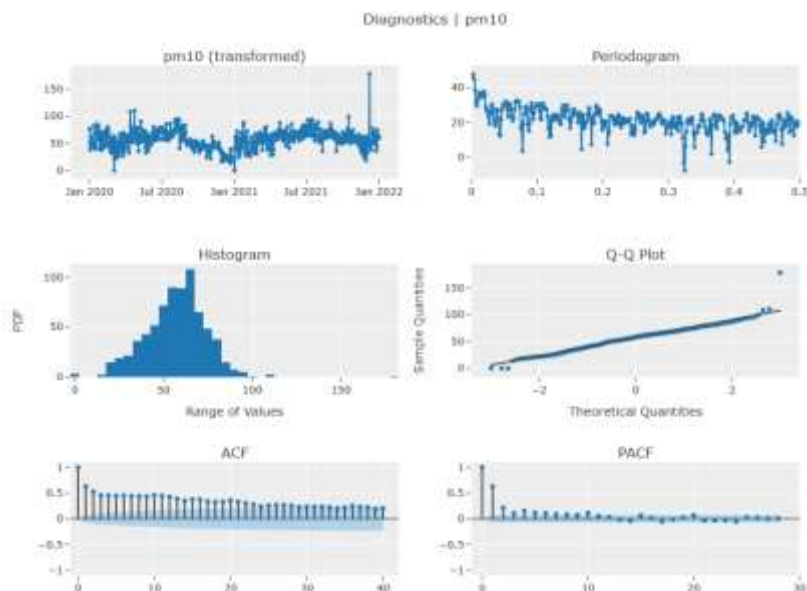
Nilai MAPE	Kriteria
<10%	Sangat Baik
10%-20%	Baik
20%-50%	Cukup
>50%	Buruk

Analisis Kinerja Model

Model *Decision Tree Regressor* yang telah melalui proses tuning dan pengujian akan digunakan untuk melakukan prakiraan indeks konsentrasi Partikulat PM10 pada satu bulan kedepan. Hasil yang didapatkan berupa grafik timeseries prediksi, dan grafik komponen. Penyusunan *Decision Tree Regressor* (DTR) melibatkan beberapa tahapan yang harus diikuti. Pertama, langkah awal adalah mengumpulkan dataset yang sesuai, memastikan adanya variabel dependen (*output*) dan variabel independen (*input*) yang relevan untuk prediksi. Setelah itu, pemilihan variabel dilakukan dengan cermat untuk memilih input yang memiliki potensi memberikan informasi signifikan terkait variabel dependen. Proses selanjutnya melibatkan pengolahan data, seperti penanganan nilai yang hilang, deteksi dan penanganan outlier, serta normalisasi atau standarisasi data. Setelah data dipersiapkan, dataset dibagi menjadi data pelatihan dan data pengujian. Data pelatihan digunakan untuk melatih model, sementara data pengujian digunakan untuk menguji kinerja model.

Pada tahap pembangunan model DTR, algoritma *Decision Tree Regression* diterapkan untuk membentuk pohon keputusan. Penentuan parameter, seperti kedalaman pohon dan kriteria pemisahan, menjadi langkah kritis untuk mengoptimalkan kinerja dan kompleksitas model. Model kemudian dilatih menggunakan data pelatihan, dan evaluasi kinerja dilakukan menggunakan data pengujian dengan metrik seperti RMSE, MAE, dan MAPE. Opsionalnya, model dapat dioptimalkan melalui penyesuaian parameter atau teknik pruning untuk mencegah overfitting. Setelah dianggap memadai, model dapat digunakan untuk membuat prediksi pada data baru. Validasi model, melalui teknik seperti cross-validation, dapat digunakan untuk memastikan kinerja yang konsisten pada dataset yang berbeda. Terakhir, dokumentasi dan interpretasi model perlu dilakukan untuk menjelaskan bagaimana variabel-variabel memengaruhi prediksi, memberikan pemahaman yang lebih mendalam terkait hasil yang diperoleh dari pohon keputusan yang telah dibangun.

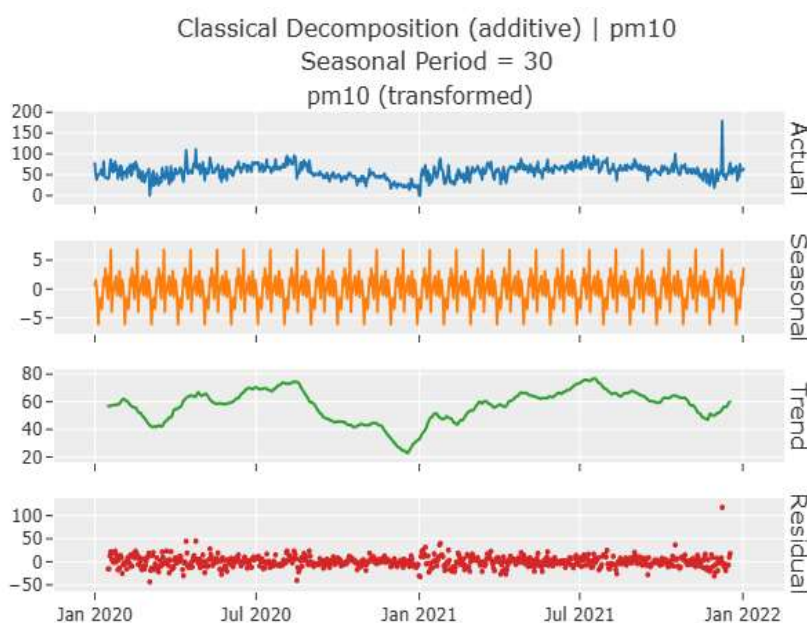
Hasil dan Pembahasan



Gambar 1. Panel Diagnostik untuk data deret waktu PM10 (*memperlihatkan fluktuasi nilai seiring berjalannya waktu (kiri atas), periodogram untuk menunjukkan komponen frekuensi (kanan atas), histogram yang menggambarkan distribusi nilai PM10 (tengah kiri), plot Q-Q yang membandingkan distribusi nilai PM10 dengan distribusi normal (tengah kanan), serta fungsi autokorelasi (ACF) (kiri bawah) dan autokorelasi parsial (PACF) (kanan bawah) untuk mendeteksi korelasi dalam data*)

Plot deret waktu yang diubah pada panel diagnostik Gambar 1 menunjukkan konsentrasi partikel halus (PM10) dari waktu ke waktu dengan variasi yang tampak mengikuti pola atau musiman, tanpa tren naik atau turun jangka panjang yang jelas, menunjukkan bahwa data umumnya statis dibandingkan dengan rata-rata. Pada saat yang sama, periodogram menunjukkan adanya siklus dominan, yang mungkin berhubungan dengan pengaruh musiman atau pola reguler lainnya dalam data. Histogram dari nilai PM10 terlihat terdistribusi tidak merata pada range 50 - 100 yang melonjak tinggi yang menandakan luasnya distribusi PM10 dengan value yang tidak merata. Q-Q plot membandingkan distribusi PM10 dengan distribusi normalnya. Terlihat titik titiknya mengikuti garis dengan sangat baik.

Fungsi autokorelasi (ACF) tidak menunjukkan korelasi yang signifikan antara nilai deret waktu dan nilai lag, yang ditunjukkan dengan penurunan nilai ACF yang cepat, dan berada dalam batas kepercayaan untuk semua interval waktu. Hal ini menunjukkan bahwa tidak terdapat masalah autokorelasi yang signifikan pada data. Sebaliknya, fungsi autokorelasi parsial (PACF) menunjukkan jeda setelah lag pertama, yang mungkin menunjukkan bahwa model *Decision Tree Regressor* mungkin merupakan pilihan yang baik untuk deret waktu ini. Singkatnya, kurva diagnostik menunjukkan bahwa data PM10 setelah hasil transformasi menunjukkan karakteristik data deret waktu yang stasioner dengan beberapa kemungkinan komponen musiman. Data ini tampaknya mendekati distribusi normal dan tidak memiliki masalah autokorelasi signifikan yang harus diselesaikan sebelum analisis atau pemodelan lebih lanjut.



Gambar 2. Decomposisi Klasik (additif) data deret waktu PM10 (Grafik ini memaparkan pemisahan data PM10 yang telah ditransformasi menjadi tiga komponen utama - nilai aktual (atas), musiman (tengah), dan tren (bawah), dilengkapi dengan residu untuk mengidentifikasi pola yang tidak dijelaskan oleh musiman dan tren pada periode Januari 2021 hingga Januari 2022)

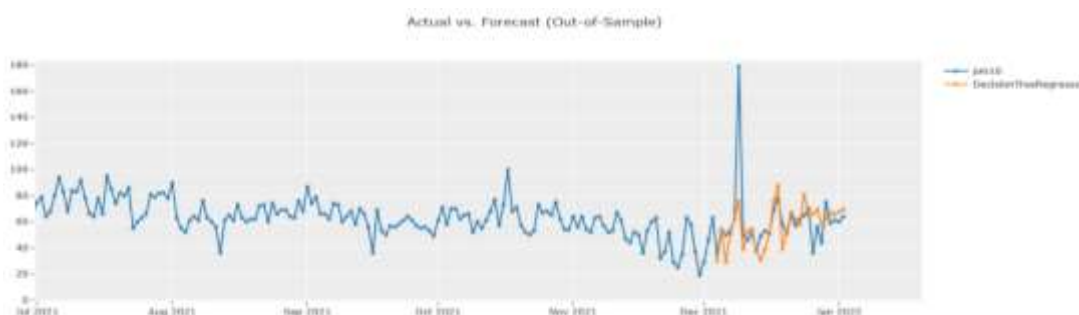
Gambar 2 mengilustrasikan dekomposisi klasik pada grafik ini membantu memahami struktur dalam data deret waktu PM10 yang telah ditransformasi. Pada bagian atas, kita melihat nilai aktual yang menunjukkan fluktuasi konsentrasi PM10 sepanjang waktu. Terlihat ada pola yang jelas dan menonjol yang bisa diidentifikasi secara visual pada bulan Desember 2021, sedangkan pada waktu yang pola yang terlihat tidak terlalu menonjol, mengindikasikan bahwa data mungkin tidak memiliki komponen musiman yang kuat, sesuai dengan periode musiman yang ditetapkan sebesar 30. pada grafik kedua menunjukkan komponen musiman

yang telah diekstrak dari data, terlihat pola yang berulang, Pola musiman yang berulang dalam konteks prediksi dapat mencerminkan fluktuasi berkala dalam data target yang sedang diprediksi. model biasanya dirancang untuk memodelkan dan memprediksi data deret waktu dengan memperhatikan pola musiman, tren, dan efek liburan. Komponen tren ditunjukkan pada bagian bawah dengan garis hijau. Terlihat sedikit variasi dalam tren ini dimana Prediksi PM10 menggunakan *Decision Tree Regressor* memperlihatkan adanya tren naik atau turun dalam jangka waktu tersebut. Grafik terakhir, residu yang ditunjukkan pada grafik terbawah, sudah ideal karena seharusnya memang acak jika semua informasi berguna telah berhasil diekstrak oleh komponen musiman dan tren. Namun, residu tidak sepenuhnya acak, ada beberapa struktur yang tersisa yang mungkin menunjukkan informasi tambahan yang tidak ditangkap oleh dekomposisi.

Tabel 4. Evaluasi Model Terhadap Data Training dan Data Testing

No	Parameter Evaluasi	Nilai	
		Data Training	Data Testing
1	RMSE	5.4167	7.5970
2	MAE	3.6370	5.4514
3	MAPE	0.2857	0.3181

Tabel 4 menjelaskan evaluasi kinerja model yang dilakukan terhadap data latih, terlihat bahwa model tersebut menunjukkan performa yang memuaskan dengan indikator-indikator statistik yang kuat. Nilai *Mean Absolute Percentage Error* (MAPE) sebesar mengindikasikan bahwa rata-rata, prediksi model menyimpang dari nilai aktual sebesar kurang lebih 28.57% yang termasuk pada tingkat kriteria error yang cukup. suatu tingkat kesalahan yang mungkin dianggap akseptabel tergantung pada konteks aplikasi model ini. Sementara itu, *Root Mean Square Error* (RMSE) mencatat nilai 5.4167. *Mean Absolute Error* (MAE) yang tercatat sebesar 3.6370 menunjukkan bahwa, tanpa mempertimbangkan arah kesalahan (*overestimate* atau *underestimate*), model memiliki rata-rata kesalahan absolut dalam jumlah tersebut ketika memprediksi data latih. Diantara ketiga perbandingan ini, yang paling baik adalah dengan metode MAE karena yang paling mendekati nol.



Gambar 3. Perbandingan Data Aktual vs Prediksi menggunakan DTR (*Decision Tree Regressor*) untuk PM10 dari tahun 2021 hingga awal 2022

Tabel 4 mendeskripsikan performa dari model ketika digunakan pada data percobaan, yang menampilkan bagaimana model generalisasi pada data sebelumnya. MAPE dengan nilai 31.81% menandakan bahwa model yang memprediksi memiliki nilai kesalahan sebanyak 31.81% dari nilai aslinya. RMSE dengan nilai kesalahan yang besar di angka 7.5970 menandakan bahwa model memiliki variabilitas yang signifikan dalam prediksi. Hal ini bisa

disebabkan oleh adanya noise dalam data dan fitur yang tidak relevan atau kurang. MAE dengan nilai 5.4514 memberikan gambaran yang lebih jelas tentang kesalahan perkiraan rata-rata model, tidak terpengaruh oleh outlier. Hal ini juga menunjukkan bahwa pada skala kesalahan absolut, model menyimpang dari nilai sebenarnya dengan rata-rata sekitar 5.4514 unit dari nilai sebenarnya.

Kesimpulan

PM10 merupakan kelompok partikulat berukuran kecil 0-10 μm , partikulat yang kecil-kecil ini merupakan risiko kesehatan yang terbesar diantara berbagai ukuran partikulat karena terhirup masuk melalui saluran pernapasan sampai dengan saluran pernapasan bagian bawah dan dideposit dalam di paru-paru. Dampak Kesehatan yang dihasilkan yang disebabkan oleh tingginya sebaran pm10 adalah kematian dini, asma, ISPA, penyakit saluran pernafasan dan bronkitis kronis. Dalam penelitian kali ini, kami mengambil wilayah DKI Jakarta sebagai wilayah percobaan. Model yang digunakan adalah *Decision Tree Regression* dengan basis *time series* untuk mem-plotting datanya. Persiapan dataset dilakukan dengan mengunduh data dari laman Indeks Standar Pencemaran Udara (ISPU) DKI Jakarta kemudian diolah menggunakan Microsoft Excel. Selanjutnya dilakukanlah pembangunan model dan uji kesalahan terhadap hasil yang sudah didapat dari implementasi *Decision Tree Regression*. Uji kesalahan dilakukan menggunakan MAE, RMSE, dan MAPE. MAPE menjadi fokus utama evaluasi. MAE digunakan untuk menghitung nilai rata-rata antara nilai aktual dalam seri waktu dan nilai yang diprediksi. MAPE menunjukkan seberapa besar kesalahan dalam meramal yang dibandingkan dengan nilai nyata dalam deret. RMSE merupakan suatu jenis kesalahan yang tergantung pada skala dan tidak dapat dibandingkan secara langsung antar skala yang berbeda. Hasil pemodelan *Decision Tree Regressor* pada parameter PM10 didapatkan nilai RMSE sebesar 7.5970, MAE sebesar 5.4514, dan MAPE sebesar 0.2857. Berbagai grafik hasil pengujian *Decision Tree Regressor* seperti PM10, grafik komponen, dan perbandingan *Actual vs Forecast* dianalisis untuk mendapatkan pemahaman yang lebih baik. Dari hasil yang didapatkan bisa disimpulkan bahwa model *Decision Tree Regressor* cukup baik dalam memprediksi PM10 di DKI Jakarta. Rekomendasi dan saran yang dapat diberikan untuk penelitian berikutnya adalah dengan menggabungkan beberapa model machine learning untuk menghasilkan model baru yang memiliki performa lebih baik. Dalam konteks perbandingan model, direkomendasikan untuk menguji performa model DTR dengan membandingkannya dengan model regresi lainnya, seperti regresi linier, guna mendapatkan pemahaman yang lebih baik mengenai keunggulan dan kelemahan model yang diusulkan. Tentu dengan hasil yang lebih baik menciptakan prediksi yang lebih akurat.

REFERENSI

- [1] Utami Rizky Putri. (2018). Skripsi "Dinamika Kekhususan Provinsi Dki Jakarta Sebagai Ibukota Negara"
- [2] Nur, Erdi, Basuki Ario Seno, Rahmi Hidayanti. (2021). Risiko Gangguan Kesehatan Masyarakat Akibat Paparan PM10 di Kota Padang. *Jurnal Kesehatan Lingkungan Indonesia*, 20 (2)
- [3] Mursinto, Djoko, Deni Kusumawardani. (2016). Estimasi Dampak Ekonomi Dari Pencemaran Udara Terhadap Kesehatan Di Indonesia. *Jurnal Kesehatan Masyarakat*, 11(2)
- [4] Hidayat A, Inaku R, Novianus C, Studi P, Masyarakat K, Kesehatan II, et al. Pengaruh Pencemaran Udara PM 2,5 dan PM 10 Terhadap Keluhan Pernapasan Anak di Ruang Terbuka Anak di DKI Jakarta The Effect of PM 2.5 and PM 10 Air pollution on Complaints of Children's Respiration in Children's Open Space in DKI Jakarta.

- [5] Oktavia A, Hustinawati. (2021). Prediksi Rata-Rata Zat Berbahaya Di Dki Jakarta Berdasarkan Indeks Standar Pencemar Udara Menggunakan Metode Long Short-Term Memory. *Jurnal Ilmiah Informatika Komputer*, 26(21)
- [6] Cahyadi, Wiji, Basir Achmad, Eko Suhartono, Fakhur Razie. (2016). Pengaruh Faktor Meteorologis Dan Konsentrasi Partikulat (PM10) Terhadap Kejadian Infeksi Saluran Pernapasan Akut (ISPA) : Studi Kasus Kecamatan Banjarbaru Selatan, Kota Banjarbaru Tahun 2014-2015, 12(3), p-ISSN 1978-8096, e-ISSN 2302-3708
- [7] Wulandari, Astri, Yusniar Hanani D, Mursid Raharjo. (2016). Analisis Risiko Kesehatan Lingkungan Paparan *Particulate Matter (PM10)* Pada Pedagang Kaki Lima Akibat Aktivitas Transportasi : Studi Kasus Jalan Kaligawe Kota Semarang, *Jurnal Kesehatan Masyarakat*, 4(3), ISSN: 2356-3346
- [8] Kurniawan, Agusta. (2017). Pengukuran Parameter Kualitas Udara (CO, NO2, SO2, O3, dan PM10) Di Bukit Kototabang Berbasis ISPU. *Jurnal Teknosains*. 7(1), p-ISSN 2089-6131, e-ISSN 2443-1311
- [9] Akuba, Manda N, Fernandez, Samuel P, Maskim, Fransisco, Ganimel, Dosso SA. (2023) Perbandingan Model Dalam Memprediksi Karbon Monoksida Di Kota Medan Bulan Januari 2023. *DijITAC 2022*, 2(2)
- [10] Thabibi Althof, Supriyatno Raden. (2023) Perbandingan Model Multiple Linear Regression Dan Decision Tree Regression (Studi Kasus: Prediksi Harga Saham Telkom, Indosat, Dan Xi. *Jurnal Ilmiah Teknologi dan Rekayasa*, 28(3)
- [11] Risanti, Indirasari, W, Suhendar, Haris. Analisis Model Prediksi Cuaca Menggunakan Support Vector Machine, Gradient Boosting, Random Forest, Dan Decision Tree. Prosiding Seminar Nasional Fisika, 2024, Vol XII, p-ISSN: 2339-0654, e-ISSN: 2476-9398
- [12] Azan Kanezar MS, Muhammad Aditya Sasmita, Agung Hari Saputra. (2023). Prediksi *Particulate Matter (PM 2.5)* di DKI Jakarta Menggunakan XGBoost. *Jurnal Meteorologi*, 2(1), ISSN: 2829-7458
- [13] Putro B, Tanzil Furqon M, Wijoyo SH. Prediksi Jumlah Kebutuhan Pemakaian Air Menggunakan Metode Exponential Smoothing (Studi Kasus: PDAM Kota Malang) [Internet]. Vol. 2. 2018. Available from: <http://j-ptiik.ub.ac.id>
- [14] Andriansyah, D, Astuti, Diah N. (2023). Perbandingan Model Prediksi Radiasi Matahari Berbasis Mesin Pembelajaran Pada Stasiun Meteorologi Fatmawati Soekarno Bengkulu. 14(1), p-ISSN 2086-5589, e-ISSN : 2763-2239
- [15] Robial S. (2018). Perbandingan Model Statistik Pada Analisis Metode Peramalan Time Series (Studi Kasus: Pt. Telekomunikasi Indonesia, Tbk Kandatel Sukabumi). *Jurnal Ilmiah SANTIKA*, 8(2)
- [16] Latifah, Retnani, Emi S.W., Priadhana E.K. (2019). Model *Decision Tree* untuk Prediksi Jadwal Kerja menggunakan *Skicit-Learn*. Seminar Nasional dan Sains 2019, p-ISSN : 2407-1846, e-ISSN: 2460-8416
- [17] Khaira, Ulfa, Norman Syarief, Zalman, Isra Hayati. (2020). Prediksi Tingkat Fertilitas Pria Dengan Algoritma Pohon Keputusan Cart. *Jurnal Ilmiah Umum Dan Kesehatan Aisyiyah*, 5(1)
- [18] Wanto, Anjar, Agus P. W. (2017). Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan Metode Backpropagation. *Jurnal & Penelitian Teknik Informatika*, 2(2), e-ISSN: 2541-2019, p-ISSN: 2541-044X
- [19] Asmoro, Achmad S. B., Wahyu S. G. I., Utomo P. (2018). Perbandingan Kinerja Hasil Seleksi Fitur pada Prediksi Kinerja Akademik Siswa Berbasis Pohon Keputusan. *Jurnal Edukasi dan Penelitian Informatika*, 4(2), e-ISSN: 2548-9364, p-ISSN: 2460-0741
- [20] Syech, Riad, Usman Malik, Riska Fitriani. (2017). Analisis Pengaruh Partikulat Matter Pm10 Terhadap Suhu, Kelembaban Udara Dan Kecepatan Angin Di Daerah Kulim Kota Pekanbaru. *Jurnal Komunikasi Fisika Indonesia*, 14(2), p-ISSN: 1412-2960, e-ISSN: 2579-521X