

Perbandingan Algoritma Regresi Linier Berganda dan Random Forest Regression untuk Prediksi Konsentrasi Particulate Matter 2,5 (PM2,5) Kota Jambi

Dinda Aulia¹, Jasmir², Sharipuddin³

¹ Program Studi Magister Sistem Informasi, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: ¹dindaul26@gmail.com, ²ijay_jasmir@yahoo.com, ³sharifbuhaira@gmail.com

Submitted :
26 Maret 2025

Revision :
24 Mei 2025

Accepted:
25 September 2025

Published:
30 September 2025

Abstrak—Permasalahan peningkatan polusi udara yang terjadi secara global juga terjadi di wilayah Indonesia termasuk di Provinsi Jambi. Polusi udara memiliki partikel-partikel debu didalamnya, salah satunya Particulate Matter (PM2,5). Konsentrasi PM2,5 di udara dipengaruhi oleh kondisi meteorologi suatu daerah, serta kejadian di sekitarnya baik itu kejadian alami ataupun kejadian yang disebabkan oleh aktivitas manusia. Penelitian ini melakukan prediksi konsentrasi PM2,5 kota Jambi dengan menggunakan algoritma Regresi Linier Berganda dan Random Forest Regression dengan suhu udara, kelembapan udara, kecepatan angin, curah hujan dan titik panas sebagai variabel bebasnya. Dalam prosesnya penelitian ini membandingkan kedua algoritma tersebut dan menilai akurasi dari masing-masing algoritma. Algoritma Regresi Linier Berganda mampu menghasilkan model yang dapat menggambarkan hubungan antara suhu udara, kelembapan udara, kecepatan angin, curah hujan dan titik panas terhadap konsentrasi PM2,5 meskipun nilai kesalahannya lebih besar dibandingkan dengan algoritma Random Forest Regression. Algoritma Random Forest Regression menghasilkan model dengan RMSE 0,148 μ gram/mm³ lebih kecil dibandingkan algoritma Regresi Linier Berganda. Pada pengujian akurasi dengan MAPE, algoritma Random Forest Regression memiliki nilai 74,0% dimana Regresi Linier Berganda memiliki nilai 73,0%, sehingga algoritma Random Forest Regression memiliki akurasi yang lebih tinggi dalam melakukan prediksi konsentrasi PM2,5.

Kata Kunci: Particulate Matter 2,5 (PM2,5), Data Mining, Prediksi, Regresi Linier Berganda dan Random Forest Regression

Abstract— The problem of increasing air pollution that occurs globally also occurs in Indonesia, including in Jambi Province. Air pollution has dust particles in it, one of which is Particulate Matter (PM2.5). The concentration of PM2.5 in the air is influenced by the meteorological conditions of an area, as well as the events around it, whether it is a natural event or an event caused by human activities. This study predicted PM2.5 concentration in Jambi city using Multiple Linear Regression and Random Forest Regression algorithms with air temperature, air humidity, wind speed, rainfall and hot spots as independent variables. In the process, this study compares the two algorithms and assesses the accuracy of each algorithm. The Multiple Linear Regression algorithm is able to generate a model that can describe the relationship between air temperature, air humidity, wind speed, rainfall and hot spots to PM2.5 concentration, although the error value is larger than that of the Random Forest Regression algorithm. The Random Forest Regression algorithm produces a model with an RMSE of 0.033 μ g/mm³ smaller than the Multiple Linear Regression algorithm. In the accuracy test with MAPE, the Random Forest Regression algorithm has a value of 74,0% where Multiple Linear Regression has a value of 73,0% so that the Random Forest Regression algorithm has a higher accuracy in predicting PM2.5 concentrations.

Keywords: Particulate Matter 2,5 (PM2,5), Data Mining, Prediction, Multiple Linear Regression and Random Forest Regression

1. PENDAHULUAN

Ancaman lingkungan terbesar terhadap kesehatan manusia yang sejalan dengan perubahan iklim yang saat ini sedang terjadi yaitu polusi udara. Paparan polusi udara setiap tahunnya diperkirakan menyebabkan tujuh juta kematian prematur dan mengakibatkan hilangnya jutaan tahun hidup sehat. [1] Permasalahan peningkatan polusi udara ini juga terjadi di wilayah Indonesia termasuk di Provinsi Jambi. Peningkatan polusi udara di Provinsi Jambi dipicu oleh berbagai hal. Salah satu penyumbang terbesar polusi udara di Provinsi Jambi yaitu berasal dari asap yang terjadi akibat kebakaran hutan dan lahan (Karhutla). Karhutla merupakan jenis bencana alam yang terjadi berulang setiap tahunnya di Provinsi Jambi. Penyebab utama karhutla di Provinsi Jambi yaitu adanya kegiatan pembakaran hutan dan lahan yang dilakukan oleh manusia untuk membuka lahan perkebunan. Keadaan cuaca yang kering pada musim kemarau memudahkan hutan dan lahan untuk terbakar karena memiliki kandungan air yang menipis. Karhutla yang terjadi di Provinsi Jambi pada tahun 2015 menyebabkan 19 orang meninggal dan 500 ribu jiwa mengalami infeksi saluran pernapasan akut (ISPA). [2] Polusi udara memiliki partikel-partikel debu didalamnya, salah satunya Particulate Matter (PM2,5).

United States of Environmental Protection Agency (USEPA) menetapkan istilah PM2,5 untuk partikel (PM) di udara yang berdiameter kurang dari 2,5 μ m. [3] PM2,5 merupakan komponen utama pada polusi udara yang terdiri dari partikel dengan diameter kurang dari 2,5 μ m. [4] PM2,5 digambarkan sebagai partikel halus yang menyebar dalam area yang luas di udara. [5] Konsentrasi PM2,5 di udara dipengaruhi oleh kondisi meteorologi

suatu daerah, serta kejadian di sekitarnya baik itu kejadian alami ataupun kejadian yang disebabkan oleh aktivitas manusia. Pengamatan konsentrasi PM_{2,5} saat ini menjadi hal yang penting mengingat banyaknya polutan di udara yang kerap mengganggu kesehatan manusia serta menimbulkan kerugian baik secara materi ataupun kerugian terhadap kerusakan lingkungan. Upaya untuk mengurangi dampak yang muncul akibat tingginya konsentrasi PM_{2,5} baik di bidang kesehatan maupun ekonomi menjadi suatu langkah penting, khususnya untuk daerah Kota Jambi yang memiliki jumlah penduduk terbanyak di Provinsi Jambi yaitu sebanyak 635,1 juta jiwa menurut BPS pada tahun 2024. [6] Prediksi konsentrasi PM_{2,5} dibutuhkan untuk membantu mengurangi dampak bagi pihak-pihak terkait.

Data mining merupakan studi atau teknik pengumpulan, pembersihan, pemrosesan dan analisis data yang dilakukan untuk mendapatkan pola serta informasi yang lebih berguna dari sekumpulan besar data. Dalam konteks prediksi konsentrasi PM_{2,5}, data mining dapat digunakan untuk menganalisis data historis PM_{2,5} dan faktor-faktor yang mempengaruhinya, sehingga dapat memberikan gambaran yang lebih jelas tentang tren dan pola yang ada. Prediksi adalah memprakirakan apa yang akan terjadi di masa yang akan datang, sehingga hasil dari prediksi erat kaitannya untuk dapat menjadi sebuah acuan dalam mengambil sebuah keputusan. [7] Terdapat beberapa metode yang biasa digunakan dalam data mining untuk tujuan prediksi seperti Regresi Linier Berganda dan Random Forest Regression.

Metode Regresi Linier Berganda merupakan metode statistik yang menggambarkan hubungan antara variabel independen dengan variabel dependen dengan sebuah persamaan yang memiliki koefisien regresi untuk masing-masing variabel dan sejumlah variabel independen. [8] Regresi Linier merupakan salah satu metode untuk prediksi yang memanfaatkan garis lurus untuk menggambarkan hubungan pada dua variabel atau lebih. [9] Variabel atau dikenal juga sebagai atribut adalah besaran yang berubah-ubah nilainya. Selanjutnya variabel tersebut terbagi atas dua jenis yaitu variabel independen atau variabel bebas dan variabel dependen atau variabel tetap. [10] Beberapa penelitian dengan tujuan prediksi yang menggunakan metode Regresi Linier Berganda sudah dilakukan. Penelitian untuk analisa curah hujan menggunakan metode Regresi Linier Berganda sebagai metode prediksi menunjukkan hasil perhitungan kesalahan sebesar 0,009 karena kesalahan yang terjadi berkisar antara 0,0-0,1 sehingga nilai kesalahan tergolong kecil. [11]

Penelitian dengan tujuan prediksi dengan cara lain juga telah dilakukan, yaitu dengan menggunakan metode Random Forest Regression. Random Forest dikenalkan oleh Leo Breiman pada tahun 2001. [12] Random Forest Regression melakukan prediksi berdasarkan rata-rata prediksi dari nilai regresi yang dibuat oleh setiap pohon dalam susunan pohon keputusan atau decision tree dengan menggunakan beberapa data sebagai masukan. [13] Metode ini pada dasarnya menggunakan metode berbasis klasifikasi dan regresi yang ditandai dengan adanya proses agregasi pohon keputusan. [14] Random Forest adalah metode pengembangan dari Algoritma C4.5 yang menggunakan beberapa decision tree untuk melakukan pelatihan pada data menggunakan sampel individu dan setiap atribut dipecah pada pohon yang dipilih antara atribut subset yang bersifat acak. [15] Penelitian oleh Jong Hong dkk. (2020) juga dilakukan untuk mencari fitur prediktor harga rumah menggunakan metode Random Forest dan regresi OLS (Ordinary Least Squares) dengan menggunakan data dari apartemen di daerah ramai di Korea Selatan dengan kurun waktu tahun 2006 hingga 2017. Metode Random Forest Regression lebih mampu menangkap kompleksitas atau non-linieritas pasar perumahan aktual sehingga dinilai dapat menjadi pelengkap untuk model regresi OLS apabila diterapkan pada data yang besar.

Beberapa penelitian yang membandingkan masing-masing metode dengan metode lainnya untuk tujuan prediksi telah dilakukan. Penelitian yang membandingkan beberapa metode statistik untuk tujuan prediksi dilakukan untuk menemukan kelebihan dan kekurangan dari setiap aspek yang dimiliki masing-masing metode yang dipilih untuk meneliti objek atau variabel tertentu. Perbandingan metode Regresi Linier Berganda, Random Forest dan Cubist Regression untuk estimasi suhu permukaan harian pada penelitian yang dilakukan Phan Thanh Noi, Jan Degener dan Martin Kappas (2017) menyajikan berbagai percobaan untuk menilai akurasi dan kecocokan masing-masing algoritma terhadap beberapa kondisi dinamis dari sejumlah dataset yang digunakan. [16]

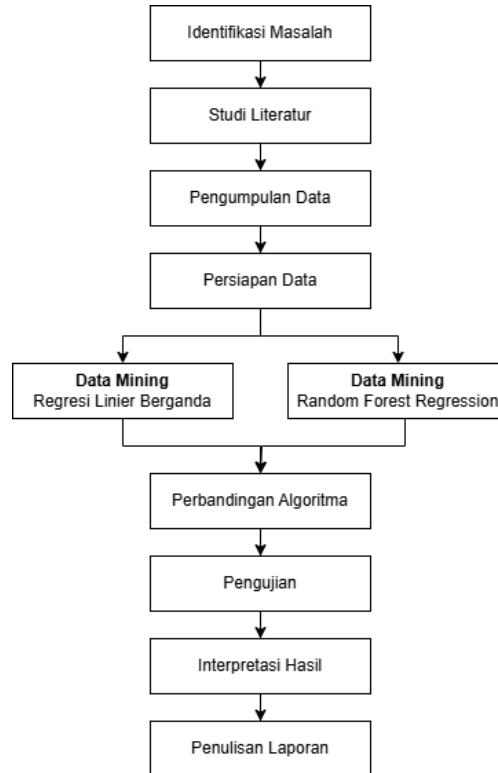
Metode Regresi Linier Berganda dan Random Forest juga dibandingkan pada penelitian Tingyuan Zhang (2023) untuk menilai seberapa mudah masing-masing model diaplikasikan dan bagaimana efisiensinya dalam hal memprediksi keuntungan penggunaan asuransi untuk beberapa golongan tertentu. Hasilnya menunjukkan metode Regresi Linier Berganda dapat menunjukkan hubungan antara masing-masing variabel independen dengan variabel dependen dan untuk metode Random Forest Regression memiliki nilai akurasi yang tinggi pada hasil prediksi. [17] Masing-masing kelebihan dari kedua metode ini dapat diuji pada model dataset lain dengan berbagai macam parameter, salah satunya pada konsentrasi PM_{2,5}.

Pentingnya mengetahui kecenderungan konsentrasi PM_{2,5} pada kondisi tertentu khususnya untuk wilayah Kota Jambi, mendorong penulis untuk melakukan penelitian yang bertujuan untuk mengetahui bagaimana metode Regresi Linier Berganda dan metode Random Forest Regression bekerja dalam melakukan prediksi terhadap nilai konsentrasi PM_{2,5}. Menggunakan beberapa parameter cuaca seperti suhu udara, kelembapan udara, kecepatan angin, curah hujan, serta jumlah titik panas pada lingkup Kota Jambi sebagai variabel independen terhadap nilai konsentrasi PM_{2,5} sebagai variabel dependen. Berdasarkan latar belakang tersebut dirumuskan beberapa masalah antara lain: 1) 1. Bagaimana perbandingan algoritma Regresi Linier Berganda dan Random Forest Regression dalam memprediksi konsentrasi PM_{2,5} di Kota Jambi? 2) 2. Bagaimana akurasi yang dihasilkan algoritma Regresi Linier Berganda dan Random Forest Regression dalam prediksi konsentrasi PM_{2,5} di Kota Jambi

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

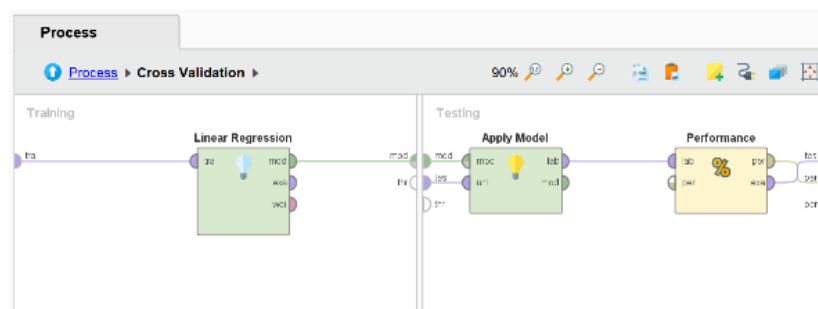
Penelitian dilakukan berdasarkan kerangka kerja penelitian seperti yang digambarkan dengan pembahasan sebagai berikut:



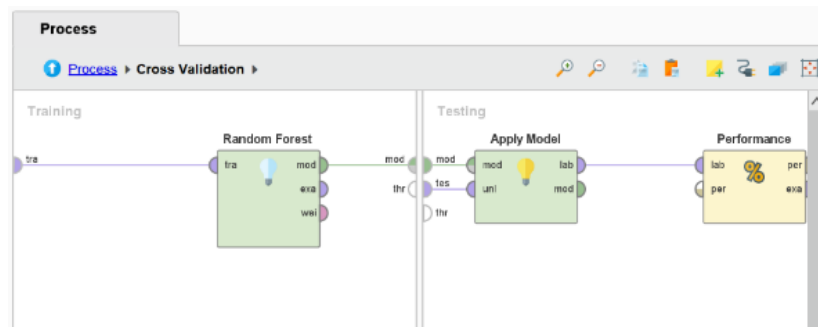
Gambar 1. Tahapan Penelitian

- a. Identifikasi Masalah. Melakukan identifikasi permasalahan pada objek yang akan diteliti, yaitu pola konsentrasi PM_{2,5} dengan unsur meteorologi lainnya. Selain itu dalam konteks pengolahan data penulis juga melakukan identifikasi pada algoritma yang digunakan pada penelitian ini dengan tujuan prediksi. Dari hasil pengamatan ini dapat disimpulkan masalah yang menjadi dasar untuk dilakukannya penelitian ini yang akan menghasilkan suatu solusi dari permasalahan tersebut.
- b. Studi Literatur. Pencarian literatur meliputi berbagai sumber online dengan tahun terbit tidak lebih dari 10 tahun yang lalu.
- c. Pengumpulan Data. Pengumpulan data untuk penelitian ini dilakukan dalam dua metode yaitu Pengamatan dan Pengumpulan data sekunder.
 1. Metode pengamatan untuk penelitian ini dilakukan dengan mempelajari beberapa unsur cuaca untuk mengetahui parameter apa saja yang mempengaruhi pola konsentrasi PM_{2,5} secara umum.
 2. Pada penelitian ini penulis menggunakan data sekunder yang didapatkan dari Stasiun Meteorologi Kelas I Sultan Thaha Jambi, berupa data sebagai berikut:
 - i. Konsentrasi PM_{2,5}, sebagai variabel tetap
 - ii. Suhu udara, sebagai variabel bebas
 - iii. Kelembapan udara, sebagai variabel bebas
 - iv. Kecepatan angin, sebagai variabel bebas
 - v. Curah hujan, sebagai variabel bebas
 - vi. Titik panas, sebagai variabel bebas
- d. Persiapan Data. Persiapan data berfungsi untuk menyesuaikan kondisi data agar dapat digunakan pada algoritma yang digunakan. Penelitian ini melakukan persiapan data menggunakan perangkat lunak Microsoft Excel dalam empat tahap yaitu:
 1. Standarisasi format. Dilakukan pemeriksaan konsistensi format file yang digunakan. Penelitian ini melakukan persiapan data pada aplikasi Microsoft Excel, sehingga file yang memiliki format csv akan diubah menjadi format Excel. Selain itu, standarisasi format juga dilakukan pada dataset yang digunakan seperti tanggal.

2. Agregasi Data. Agregasi data dilakukan pada data yang pada awalnya bersifat perjam menjadi data harian dengan menggunakan rata-rata nilai atau jumlah agar sesuai dengan sifat masing-masing parameter cuaca dalam mempengaruhi konsentrasi PM2,5. Agregasi data dilakukan dengan menggunakan PivotTable pada Microsoft Excel.
 3. Penggabungan Data. Data yang telah disiapkan sebelumnya digabungkan menjadi satu tabel dalam bentuk Excel yang selanjutnya menjadi tabel yang siap dilakukan proses selanjutnya.
 4. Penanganan Data Hilang dan *Outlier*. Data yang diambil dari Stasiun Meteorologi Kelas I Sultan Thaha Jambi adalah data yang dihasilkan langsung oleh alat pengamatan sehingga terdapat kemungkinan bahwa ada data yang hilang dikarenakan masalah teknis. Tahap ini menghilangkan seluruh data apabila terdapat satu data yang hilang pada hari tersebut. Selain itu *Outlier* atau pencilan sangat berpengaruh terhadap hasil regresi pada algoritma Regresi Linier Berganda. Tahap ini juga tahap dimana outlier dihilangkan.
- e. Data Mining. Penerapan data mining untuk penelitian ini yaitu menggunakan metode Regresi Linier Berganda dan Random Forest Regression. Pada kedua metode dilakukan pengoperasian pada aplikasi RapidMiner dengan menggunakan metode *Cross Validation*.



Gambar 2. Tampilan Proses Cross Validation untuk Algoritma Regresi Linier Berganda pada RapidMiner



Gambar 3. Tampilan Proses Cross Validation untuk Algoritma Random Forest Regression pada RapidMiner

Gambar 2 dan Gambar 3 menunjukkan operator yang digunakan pada pengoperasian algoritma Regresi Linier Berganda dan Random Forest pada aplikasi RapidMiner secara berurutan.

- f. Perbandingan Algoritma. Memberikan gambaran dan perbandingan terhadap kemampuan dari masing-masing algoritma dalam memprediksi konsentrasi PM2,5.
- g. Pengujian. Pengujian dilakukan dengan mencari metrik evaluasi yang meliputi R^2 , *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE) dan *Mean Absolute Percentage Error* (MAPE). Dikarenakan konsep regresi yang tidak dapat memberikan persentase akurasi, digunakan nilai MAPE yang nantinya akan membantu untuk mendapatkan nilai akurasi.
- h. Interpretasi Hasil. Hasil dari pemodelan oleh masing-masing algoritma yang telah dilakukan selanjutnya dianalisis dan ditampilkan bagaimana kesalahan yang dihasilkan serta nilai akurasinya terhadap prediksi konsentrasi PM2,5.
- i. Penulisan Laporan. Tahapan penelitian yang dilakukan sebelumnya selanjutnya dituangkan dalam laporan penelitian.

3. HASIL DAN PEMBAHASAN

3.1 Perbandingan Algoritma

Pada penerapan algoritma Regresi Linier Berganda dan Random Forest Regression pada RapidMiner didapatkan beberapa hal yang dapat dibandingkan:

a. Bentuk Model

1. Algoritma Regresi Linier Berganda memang dikenal dengan kemampuannya untuk menggambarkan hubungan antara variabel bebas dalam mempengaruhi variabel tetap. Dengan kemampuan ini, algoritma ini memiliki keunggulan untuk dapat menggambarkan secara langsung hubungan masing-masing variabel bebas terhadap nilai konsentrasi PM2,5. Pada hasil algoritma Regresi Linier didapatkan informasi koefisien masing-masing variabel bebas seperti ditampilkan pada Tabel 1.

Tabel 1. Nilai Koefisien hasil algoritma Regresi Linier Berganda

Atribut	Koefisien
Average of AT (°C)	-0,435
Average of RH (%)	-0,383
Wind Speed (kt)	-2,343
Sum of RR (mm)	-0,098
Hotspot	10,978

Dari Tabel 1. dapat dijelaskan nilai koefisien dari masing-masing variabel bebas terhadap nilai konsentrasi PM2,5.

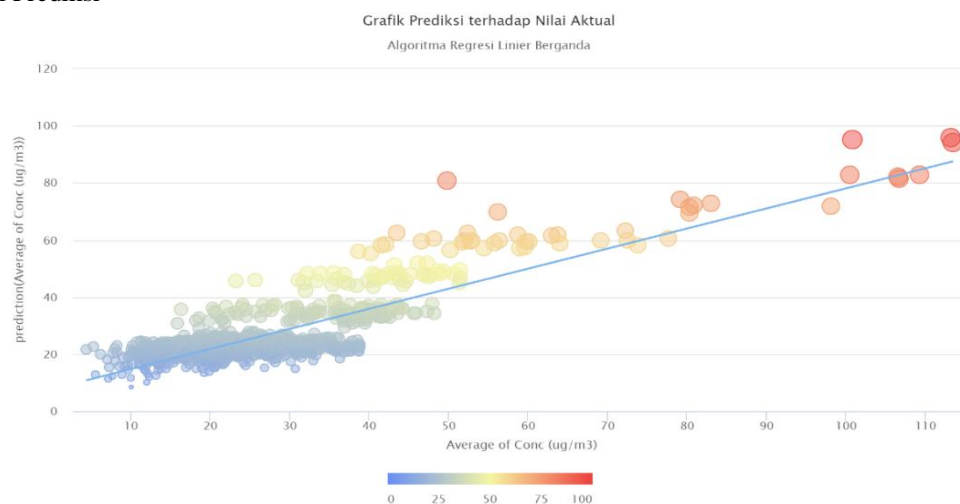
2. Random Forest Regression. Algoritma Random Forest Regression yang merupakan kumpulan dari beberapa pohon regresi tidak mengasumsikan hubungan linier pada setiap variabel, sehingga algoritma ini tidak dapat memberikan nilai koefisien yang menggambarkan pengaruh dari suhu, kelembapan udara, kecepatan angin, curah hujan dan hotspot terhadap konsentrasi PM2,5 seperti yang dilakukan oleh algoritma Regresi Linier Berganda.

Random Forest Regression lebih menilai kepentingan dari setiap fitur untuk mengevaluasi kontribusi masing-masing variabel bebas terhadap prediksinya.

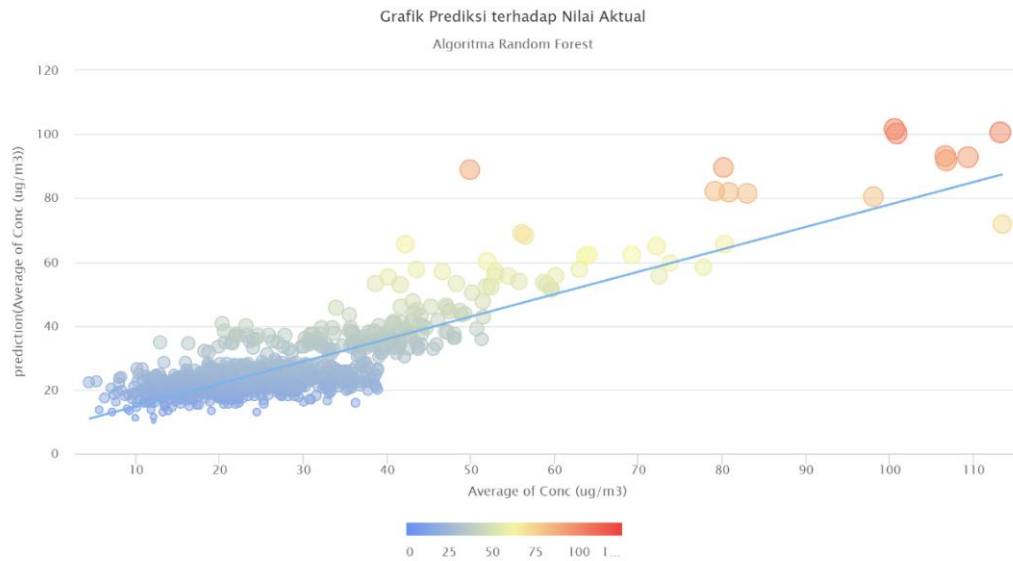
b. Kecepatan Eksekusi

1. Algoritma Regresi Linier Berganda melakukan perhitungan langsung terhadap variabel bebas dan variabel tetap sebanyak satu kali untuk setiap data. Hal ini menjadikan Regresi Linier Berganda memerlukan waktu yang lebih singkat untuk menghasilkan prediksi.
2. Algoritma Random Forest Regression cenderung membutuhkan waktu yang sedikit lebih lama dibandingkan dengan Regresi Linier Berganda. Hal ini disebabkan oleh jumlah operasi yang lebih banyak. Dalam Random Forest Regression, jumlah model yang dihasilkan setara dengan jumlah pohon. Setiap pohon harus melalui proses pembuatan, yang melibatkan pembagian data secara iteratif di setiap cabang, sehingga jumlah operasi meningkat secara signifikan.

c. Hasil Prediksi



Gambar 4. Grafik Prediksi Konsentrasi PM2.5 Terhadap Nilai Konsentrasi PM2,5 Sebenarnya Pada Regresi Linier Berganda



Gambar 5. Grafik prediksi konsentrasi PM2.5 terhadap nilai konsentrasi PM2,5 sebenarnya pada Random Forest Regression

Pada kedua grafik yang ditampilkan garis biru lurus yang menggambarkan persamaan linier yang menunjukkan model dari masing-masing algoritma. Terlihat model Random Forest Regression khususnya pada nilai menghasilkan nilai yang lebih banyak mendekati garis biru linier yang merupakan model dari algoritma tersebut. Pada kedua grafik terlihat masing-masing algoritma menghasilkan beberapa nilai prediksi yang jauh dari nilai aktual.

3.2 Hasil Pengujian

Pengujian dilakukan pada perangkat lunak RapidMiner dengan operator Performance. Pada penelitian ini pengujian dilakukan dengan membandingkan nilai R2, RMSE dan MAE serta pengujian akurasi dengan mencari nilai MAPE sehingga didapatkan hasil sebagai berikut.

Tabel 2. Hasil Pengujian

Algoritma	Regresi Linier Berganda	Random Forest Regression
R ²	0,585	0,588
RMSE (µg/m ³)	6,904	6,756
MAE (µg/m ³)	5,672	5,490
MAPE (%)	26,1	26,0
Akurasi (%)	73,9	74,0

Secara umum kedua algoritma memiliki kemampuan yang cukup baik dalam melakukan prediksi terhadap nilai konsentrasi PM2,5 di Kota Jambi. Pada nilai R2, terdapat selisih sebesar 0,003 dengan Random Forest Regression memiliki nilai yang lebih besar. Hal ini menunjukkan model yang dihasilkan Random Forest Regression mampu menjelaskan variasi data sedikit lebih baik dibandingkan dengan Regresi Linier Berganda. Pada pengujian RMSE terdapat selisih sebesar 0,148 µg/m3 dengan algoritma Random Forest Regression memiliki nilai yang lebih kecil. Hal ini menunjukkan bahwa algoritma Random Forest Regression memiliki kesalahan yang lebih kecil dalam membuat prediksi nilai konsentrasi PM2,5 dibandingkan dengan algoritma Regresi Linier Berganda.

Pada pengujian akurasi menggunakan MAPE didapatkan nilai akurasi Random Forest Regression sebesar 74,0% dan Regresi Linier Berganda sebesar 73,9%. Terdapat selisih sebesar 0,1% dengan nilai Random Forest Regression lebih besar dari Regresi Linier Berganda. Hal ini menunjukkan Random Forest Regression memiliki akurasi yang lebih baik dari Regresi Linier Berganda. Namun dengan kecilnya selisih akurasi dari masing-masing algoritma menandakan kedua algoritma sudah menghasilkan model yang cukup baik dengan kemampuan yang hampir setara. Nilai akurasi yang tidak mencapai 80% dapat disebabkan oleh data yang bersifat sangat variatif mengingat nilai konsentrasi PM2,5 yang fluktuatif karena dipengaruhi bukan hanya cuaca, namun dapat dipengaruhi juga oleh aktivitas manusia dan factor lingkungan lain. Terlihat juga nilai RMSE yang lebih besar daripada MAE, hal ini mengindikasikan bahwa terdapat beberapa prediksi dengan kesalahan yang besar.

4. KESIMPULAN

Berdasarkan hasil penelitian dan tujuan penelitian, dapat disimpulkan dua hal. Algoritma Random Forest Regression memiliki kemampuan yang lebih baik dibandingkan dengan algoritma Regresi Linier Berganda dalam memprediksi nilai konsentrasi PM_{2,5} yang ditandai dengan nilai RMSE yang lebih kecil. Pada algoritma Regresi Linier Berganda terdapat keunggulan dalam menggambarkan hubungan dari suhu udara, kelembapan udara, kecepatan angin, curah hujan dan banyaknya titik panas dalam mempengaruhi nilai konsentrasi PM_{2,5}. Hal kedua yang didapatkan dari penelitian ini yaitu akurasi yang dihasilkan oleh algoritma Random Forest Regression memiliki nilai 74,0% yang lebih besar dibandingkan dengan Regresi Linier Berganda yang memiliki nilai 73,9%, sehingga algoritma Random Forest Regression memiliki akurasi yang lebih baik dibandingkan algoritma Regresi Linier Berganda dalam melakukan prediksi terhadap konsentrasi PM_{2,5} di Kota Jambi.

REFERENCES

- [1] World Health Organization, "New WHO Global Air Quality Guidelines aim to save millions of lives from air pollution." Accessed: Oct. 25, 2024. [Online]. Available: <https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution>
- [2] A. Fauziah, M. Zuhdi, and H. Syarifuddin, "Analisis Distribusi Asap Dampak Kebakaran Hutan dan Lahan di Provinsi Jambi," *Jurnal Pembangunan Berkelanjutan*, vol. 6, no. 2, pp. 10–25, 2023, doi: 10.22437/jpb.v6i2.30262.
- [3] United States Environmental Protection Agency, "Particulate Matter (PM) Basics." Accessed: Oct. 25, 2024. [Online]. Available: <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics>
- [4] W. Xu *et al.*, "The influence of PM_{2.5} exposure on kidney diseases," Jan. 07, 2022, *SAGE Publications Ltd.* doi: 10.1177/096032712111069982.
- [5] P. Thangavel, D. Park, and Y. C. Lee, "Recent Insights into Particulate Matter (PM_{2.5})-Mediated Toxicity in Humans: An Overview," Jun. 01, 2022, *MDPI*. doi: 10.3390/ijerph19127511.
- [6] Badan Pusat Statistik Provinsi Jambi, "Penduduk Menurut Kabupaten/Kota di Provinsi Jambi." Accessed: Nov. 06, 2024. [Online]. Available: <https://jambi.bps.go.id/id/statistics-table/2/MjEwOCMy/penduduk-menurut-kabupaten-kota-di-provinsi-jambi.html>
- [7] M. Sulistiyono, B. Satria, A. Sidauruk, and R. Wardhana, "RAINFALL PREDICTION USING MULTIPLE LINEAR REGRESSION ALGORITHM," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 9, no. 1, pp. 17–22, Aug. 2023, doi: 10.33480/jitk.v9i1.4203.
- [8] A. Lin, "Walmart Sales Prediction Using Multiple Linear Regression," 2024.
- [9] P. Sari Ramadhan and N. Safitri STMIK Triguna Dharma, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang," vol. 18, no. SAINTIKOM, pp. 55–61, 2019, [Online]. Available: <https://sirusa.bps.go.id/index.php>
- [10] D. Tampubolon and D. Saripurna, "Implementasi Regresi Linier Berganda Untuk Memprediksi Tingkat Penjualan Alat Kelistrikan," *Jurnal CyberTech*, vol. 3, no. 1, pp. 176–185, 2020, [Online]. Available: <https://ojs.trigunadharmas.ac.id/>
- [11] M. Maulita, "PENDEKATAN DATA MINING UNTUK ANALISA CURAH HUJAN MENGGUNAKAN METODE REGRESI LINEAR BERGANDA (STUDI KASUS: KABUPATEN ACEH UTARA)," 2023. [Online]. Available: <http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexMayaMaulita|http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index>
- [12] D. Pramesti and W. M. Baihaqi, "Perbandingan Prediksi Jumlah Transaksi Ojek Online Menggunakan Regresi Linier dan Random Forest," *Generation Journal*, vol. 7, Oct. 2023.
- [13] J. Hong, H. Choi, and W. S. Kim, "A house price valuation based on the random forest approach: The mass appraisal of residential property in south korea," *International Journal of Strategic Property Management*, vol. 24, no. 3, pp. 140–152, Mar. 2020, doi: 10.3846/ijspm.2020.11544.

- [14] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, vol. 1, no. 1, pp. 27–31, 2018.
- [15] R. Prasetyo *et al.*, "Komparasi Algoritma Logistic Regression dan Random Forest pada Prediksi Cacat Software," 2021. [Online]. Available: <http://promise.site.uottawa.ca>.
- [16] P. T. Noi, J. Degener, and M. Kappas, "Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data," *Remote Sens (Basel)*, vol. 9, no. 5, May 2017, doi: 10.3390/rs9050398.
- [17] T. Zhang, "Prediction for Insurance Premiums Based on Random Forest and Multiple Linear Regression," 2023.