

Pendekatan Machine Learning Berbasis Fitur Geospasial Imputasi Nilai AADT yang Hilang: Studi Kasus Texas

Afridayani^{1,*}, Afrisawati², Rizky Maulidya Afifa², Cut Try Utari³

¹ Program Studi Magister (S-2) Sains Data dan Kecerdasan Buatan, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Medan, Indonesia

² Program Studi Manajemen Informatika, Fakultas Teknik Komputer dan Informatika, Politeknik Negeri Medan, Medan, Indonesia

³ Program Studi Teknologi Rekayasa Perangkat Lunak, Fakultas Teknik Komputer dan Informatika, Politeknik Negeri Medan, Medan, Indonesia

Email: ^{1,*} afridayani@students.usu.ac.id, ² afriawati@polmed.ac.id, ³ rizkymaulidya@polmed.ac.id, ⁴ cuttryutari@polmed.ac.id

Email Penulis Korespondensi: afridayani@students.usu.ac.id

Submitted 15-04-2026; Accepted 27-04-2026; Published 30-04-2026

Abstrak

Data *Annual Average Daily Traffic* (AADT) merupakan indikator penting dalam perencanaan transportasi dan evaluasi kinerja jaringan jalan. Namun, keberadaan *missing values* akibat gangguan sensor atau ketidaktercatatan data dapat menurunkan kualitas analisis. Penelitian ini bertujuan untuk mengestimasi nilai AADT yang hilang menggunakan pendekatan *machine learning* berbasis fitur geospasial dengan studi kasus di Texas, Amerika Serikat. Data *Automatic Traffic Recorder* (ATR) diintegrasikan dengan atribut jaringan jalan dari OpenStreetMap (OSM) melalui proses *spatial join* untuk menghasilkan fitur seperti klasifikasi jalan, jumlah lajur, dan batas kecepatan. Model Random Forest digunakan untuk membangun model estimasi berdasarkan data valid (AADT > 0). Hasil evaluasi menunjukkan nilai koefisien determinasi (R^2) sebesar 0,548, yang mengindikasikan bahwa fitur geospasial mampu menjelaskan variasi AADT secara signifikan. Proses imputasi berhasil menghasilkan dataset dengan tingkat kelengkapan 100% serta pola distribusi spasial yang konsisten dengan hierarki jaringan jalan dan wilayah metropolitan. Pendekatan ini menunjukkan bahwa integrasi data spasial dan *machine learning* efektif dalam meningkatkan integritas data lalu lintas untuk mendukung pengambilan keputusan berbasis data.

Kata Kunci: Machine Learning; AADT; Imputasi Data; Geospasial; Random Forest

Abstract

Annual Average Daily Traffic (AADT) is an important indicator in transportation planning and road network performance evaluation. However, missing values due to sensor interference or unrecorded data can reduce the quality of the analysis. This study aims to estimate missing AADT values using a geospatial feature-based machine learning approach with a case study in Texas, United States. Automatic Traffic Recorder (ATR) data is integrated with road network attributes from OpenStreetMap (OSM) through a spatial join process to produce features such as road classification, number of lanes, and speed limits. A Random Forest model is used to build an estimation model based on valid data (AADT > 0). The evaluation results show a coefficient of determination (R^2) of 0.548, indicating that geospatial features can significantly explain variations in AADT. The imputation process successfully produced a dataset with a 100% convenience level and a spatial distribution pattern consistent with the road network hierarchy and metropolitan area. This approach demonstrates that the integration of spatial data and machine learning is effective in improving the integrity of traffic data to support data-driven decision making.

Keywords: Machine Learning; AADT; Data Imputation; Geospatial; Random Forest

1. PENDAHULUAN

Data lalu lintas merupakan elemen penting dalam perencanaan transportasi dan pengambilan keputusan berbasis data [1]. Salah satu indikator utama yang digunakan untuk merepresentasikan intensitas pergerakan kendaraan adalah *Annual Average Daily Traffic* (AADT), yaitu rata-rata jumlah kendaraan harian dalam satu tahun pada suatu titik pengamatan [2]. Nilai AADT umumnya diperoleh melalui stasiun *Automatic Traffic Recorder* (ATR) yang merekam volume kendaraan secara periodik [3].

Eksplorasi awal terhadap dataset ATR yang bersumber dari Kaggle menunjukkan bahwa negara bagian Texas memiliki kepadatan titik perekaman tertinggi serta nilai AADT maksimum yang signifikan dibandingkan wilayah lainnya di Amerika Serikat [4]. Oleh karena itu, penelitian ini difokuskan pada studi kasus di Texas yang memiliki kompleksitas jaringan jalan dan intensitas lalu lintas yang tinggi. Namun, hasil identifikasi data menemukan adanya anomali berupa nilai AADT sebesar 0 pada sejumlah titik yang terletak di jaringan jalan utama berdasarkan data OpenStreetMap (OSM) [5]. Secara logis, kondisi ini tidak merepresentasikan situasi lalu lintas sebenarnya dan mengindikasikan adanya *missing values* akibat gangguan sensor atau ketidaklengkapan pencatatan data [6].

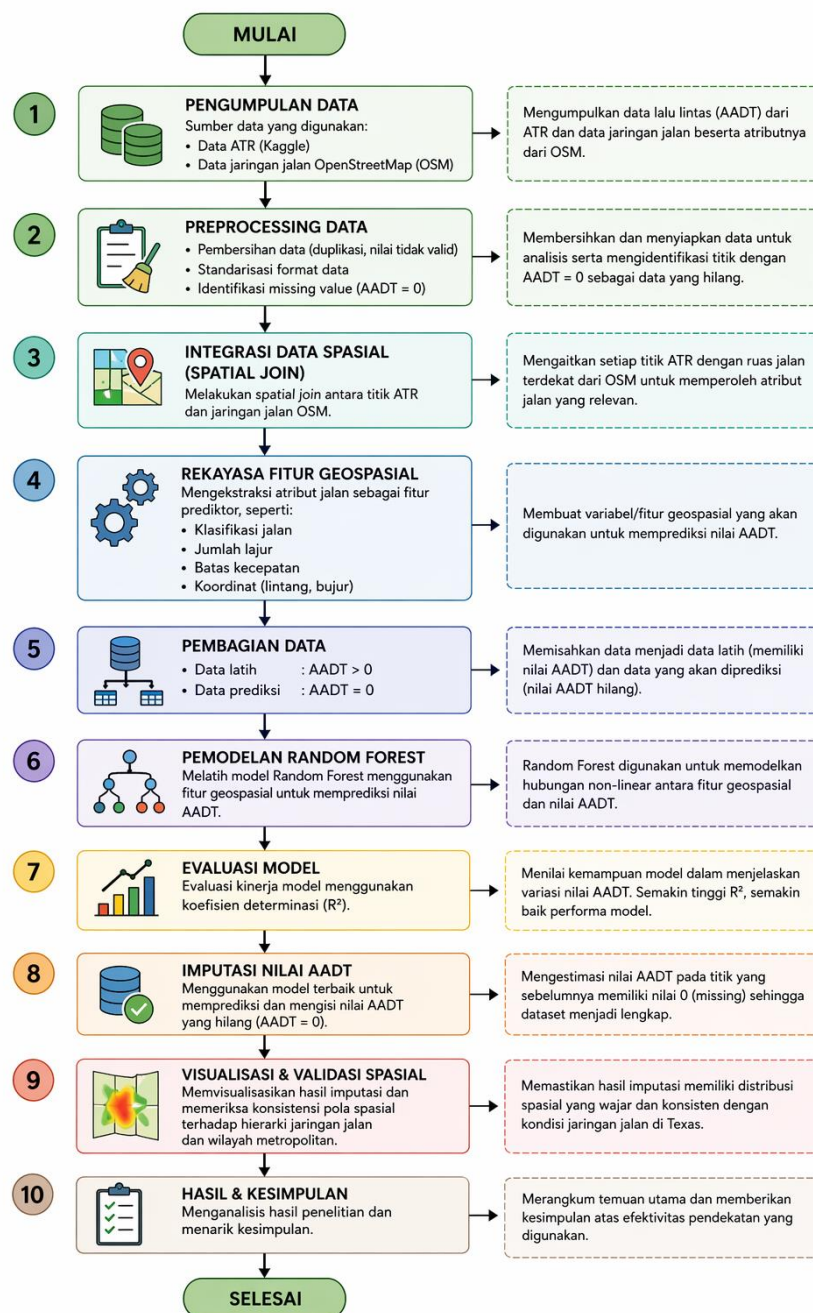
Keberadaan *missing values* pada data lalu lintas dapat menurunkan kualitas analisis dan berpotensi menghasilkan bias dalam perumusan kebijakan. Beberapa penelitian sebelumnya telah mengkaji penanganan data tidak lengkap menggunakan pendekatan *machine learning*, seperti prediksi lalu lintas berbasis data temporal dan mobilitas kendaraan [7]. Selain itu, metode seperti Random Forest juga telah digunakan dalam berbagai kasus prediksi, seperti prediksi produksi dan klasifikasi data, yang menunjukkan kemampuan dalam menangani hubungan non-linear dan data kompleks [8]. Penelitian lain juga memanfaatkan data geospasial untuk analisis jaringan jalan dan pengambilan keputusan berbasis lokasi.

Namun demikian, sebagian besar penelitian tersebut masih berfokus pada prediksi nilai atau klasifikasi berbasis data numerik dan temporal, serta belum secara spesifik mengintegrasikan atribut geospasial infrastruktur jalan dalam proses imputasi data lalu lintas. Pemanfaatan data OpenStreetMap (OSM) sebagai sumber fitur geospasial dalam mengisi nilai AADT yang hilang masih relatif terbatas. Selain itu, pendekatan imputasi konvensional umumnya belum mempertimbangkan keterkaitan antara karakteristik fisik jalan dan distribusi spasial volume kendaraan.

Berdasarkan kesenjangan tersebut, penelitian ini mengusulkan pendekatan *machine learning* berbasis fitur geospasial untuk mengestimasi nilai AADT yang hilang. Fitur prediktif diekstraksi dari atribut infrastruktur jalan OpenStreetMap dan dimodelkan menggunakan algoritma Random Forest untuk menangkap hubungan non-linear antar variabel [9]. Penelitian ini bertujuan untuk meningkatkan kelengkapan dan integritas data lalu lintas secara spasial sehingga dataset yang dihasilkan lebih representatif dan dapat digunakan sebagai dasar analisis lanjutan maupun pengambilan kebijakan transportasi berbasis data [10].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Penjelasan tahapan penelitian yaitu:

a. Pengumpulan Data

Pada tahap ini, data dikumpulkan dari dua sumber utama, yaitu data lalu lintas (*Automatic Traffic Recorder/ATR*) yang berisi nilai AADT, serta data jaringan jalan dari OpenStreetMap (OSM). Kedua data ini digunakan sebagai dasar dalam proses analisis [11].

b. Preprocessing Data

Data yang telah dikumpulkan kemudian dibersihkan dari nilai yang tidak valid atau duplikat. Pada tahap ini juga dilakukan identifikasi nilai AADT yang bernilai 0, yang dianggap sebagai *missing values* [12].

c. Integrasi Data Spasial (*Spatial Join*)

Tahap ini bertujuan untuk menggabungkan data ATR dengan data jaringan jalan OSM. Setiap titik ATR dihubungkan dengan ruas jalan terdekat sehingga diperoleh atribut jalan yang sesuai [13].

d. Rekayasa Fitur Geospasial

Pada tahap ini, atribut jalan seperti jenis jalan, jumlah lajur, batas kecepatan, dan koordinat diolah menjadi fitur yang akan digunakan dalam model *machine learning* [14].

e. Pembagian Data

Data kemudian dipisahkan menjadi dua bagian, yaitu data latih (data dengan nilai AADT tersedia) dan data prediksi (data dengan AADT = 0 yang akan diisi nilainya) [15].

f. Pemodelan *Random Forest*

Model *Random Forest* digunakan untuk mempelajari hubungan antara fitur geospasial dan nilai AADT. Model ini dilatih menggunakan data latih yang tersedia.

g. Evaluasi Model

Kinerja model dievaluasi menggunakan nilai R^2 untuk mengetahui seberapa baik model dalam menjelaskan variasi data AADT.

h. Imputasi Nilai AADT

Model yang telah dilatih digunakan untuk memprediksi dan mengisi nilai AADT yang sebelumnya hilang (*missing values*), sehingga data menjadi lengkap [16].

i. Visualisasi dan Validasi Spasial

Hasil imputasi divisualisasikan dalam bentuk peta untuk memastikan bahwa pola distribusi lalu lintas yang dihasilkan sesuai dengan kondisi nyata.

j. Hasil dan Kesimpulan

Tahap akhir adalah menganalisis hasil penelitian dan menarik kesimpulan terkait efektivitas metode yang digunakan [17].

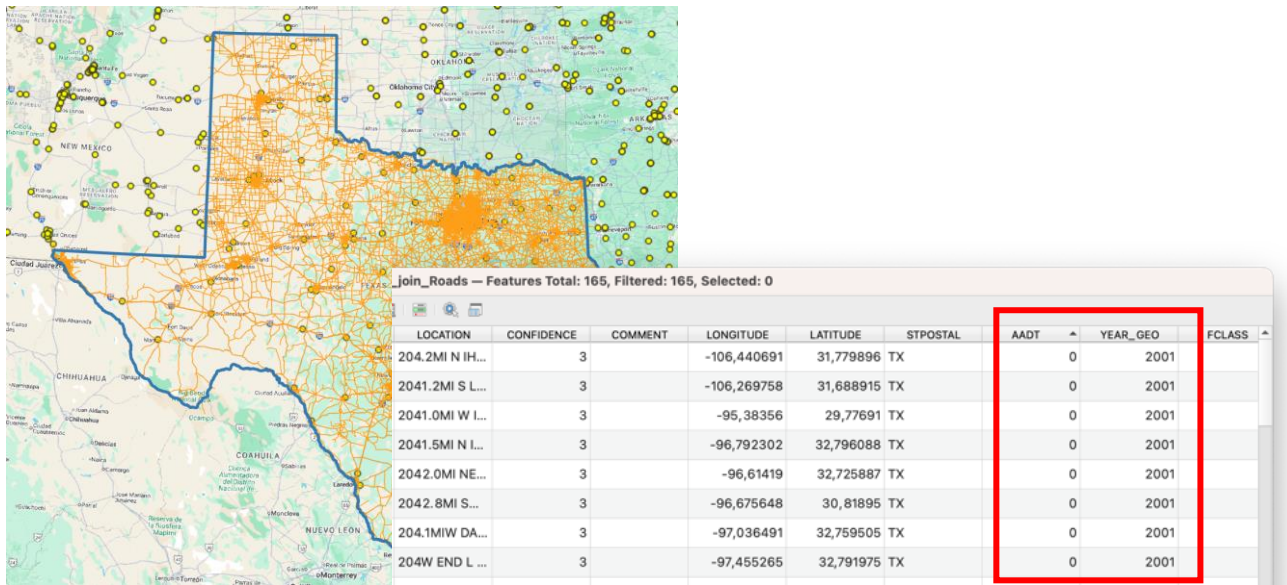
2.2 Kajian Metode *Random Forest*

Random Forest merupakan salah satu algoritma *machine learning* berbasis *ensemble learning* yang menggabungkan banyak pohon keputusan (*decision tree*) untuk meningkatkan akurasi dan mengurangi *overfitting* [18]. Algoritma ini bekerja dengan membangun beberapa pohon keputusan dari subset data yang berbeda, kemudian menghasilkan prediksi akhir berdasarkan rata-rata (untuk regresi) atau voting (untuk klasifikasi). Keunggulan *Random Forest* adalah kemampuannya dalam menangani data dengan hubungan non-linear, tahan terhadap noise, serta tidak memerlukan asumsi distribusi data tertentu [19]. Selain itu, algoritma ini juga mampu menangani data dengan jumlah fitur yang banyak, sehingga sesuai digunakan dalam penelitian ini yang memanfaatkan berbagai atribut geospasial sebagai variabel prediktor. Dalam konteks penelitian ini, *Random Forest* digunakan untuk memodelkan hubungan antara karakteristik infrastruktur jalan dan nilai AADT, serta mengestimasi nilai yang hilang berdasarkan pola yang dipelajari dari data yang tersedia [20].

3. HASIL DAN PEMBAHASAN

3.1 Identifikasi dan Pengumpulan Data

Pada tahap awal, dilakukan eksplorasi terhadap dataset *Automatic Traffic Recorder* (ATR) untuk mengidentifikasi permasalahan pada data. Hasil analisis menunjukkan adanya sejumlah titik dengan nilai AADT sebesar 0 yang berada pada ruas jalan utama. Nilai tersebut tidak sesuai dengan kondisi nyata sehingga dikategorikan sebagai *missing values*. Selanjutnya, dilakukan pembersihan data untuk memastikan tidak terdapat duplikasi atau nilai yang tidak valid. Data dengan nilai AADT > 0 digunakan sebagai data latih, sedangkan data dengan AADT = 0 disiapkan sebagai data yang akan diprediksi. Untuk memperkuat hasil identifikasi, dilakukan visualisasi distribusi awal nilai AADT pada wilayah Texas. Gambar 1 menunjukkan bahwa terdapat beberapa titik pada ruas jalan utama yang memiliki nilai AADT = 0.



Gambar 2. Peta distribusi awal AADT Texas

3.2 Proses Integrasi Data Spasial

Tahap berikutnya adalah integrasi data ATR dengan data jaringan jalan dari *OpenStreetMap* (OSM) menggunakan metode *spatial join*. Proses ini bertujuan untuk mengaitkan setiap titik ATR dengan atribut jalan terdekat.

Hasil dari proses ini adalah dataset baru yang telah dilengkapi dengan atribut geospasial seperti klasifikasi jalan, jumlah lajur, batas kecepatan, serta koordinat lokasi. Dataset inilah yang selanjutnya digunakan dalam proses pemodelan.

3.3 Rekayasa Fitur dan Pembentukan Dataset

Pada tahap ini dilakukan proses rekayasa fitur dengan memilih atribut yang relevan sebagai variabel prediktor. Fitur yang digunakan meliputi klasifikasi jalan, jumlah lajur, batas kecepatan, status satu arah, serta koordinat geografis.

Selanjutnya, dataset dibagi menjadi dua bagian, yaitu:

- Data latih (memiliki nilai AADT)
- Data prediksi (nilai AADT hilang)

Pembagian ini bertujuan untuk melatih model agar dapat mempelajari pola hubungan antara fitur geospasial dan nilai AADT.

3.4 Proses Pemodelan *Random Forest*

Model *Random Forest* dibangun menggunakan data latih yang telah dipersiapkan. Algoritma ini bekerja dengan membentuk sejumlah pohon keputusan yang kemudian digabungkan untuk menghasilkan prediksi yang lebih stabil. Selama proses pelatihan, model mempelajari hubungan antara variabel *input* (fitur geospasial) dan *output* (nilai AADT). Setelah model terbentuk, model digunakan untuk memprediksi nilai AADT pada data yang sebelumnya memiliki nilai 0.

3.5 Evaluasi Kinerja Model

Setelah proses pelatihan, dilakukan evaluasi model menggunakan koefisien determinasi (R^2). Hasil evaluasi menunjukkan nilai R^2 sebesar 0,548, yang berarti model mampu menjelaskan sekitar 54,8% variasi data. Nilai ini menunjukkan bahwa fitur geospasial memiliki pengaruh yang cukup signifikan terhadap nilai AADT, meskipun belum sepenuhnya menjelaskan seluruh variasi data. Visualisasi hasil evaluasi model ditunjukkan pada Gambar 2.

```
# Split data untuk validasi akurasi
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

# Inisialisasi Model
model = RandomForestRegressor(n_estimators=100, random_state=42)

# Latih Model
model.fit(X_train, y_train)

# Cek Akurasi
print(model.score(X_test, y_test))

... 0.5482771453742155
```

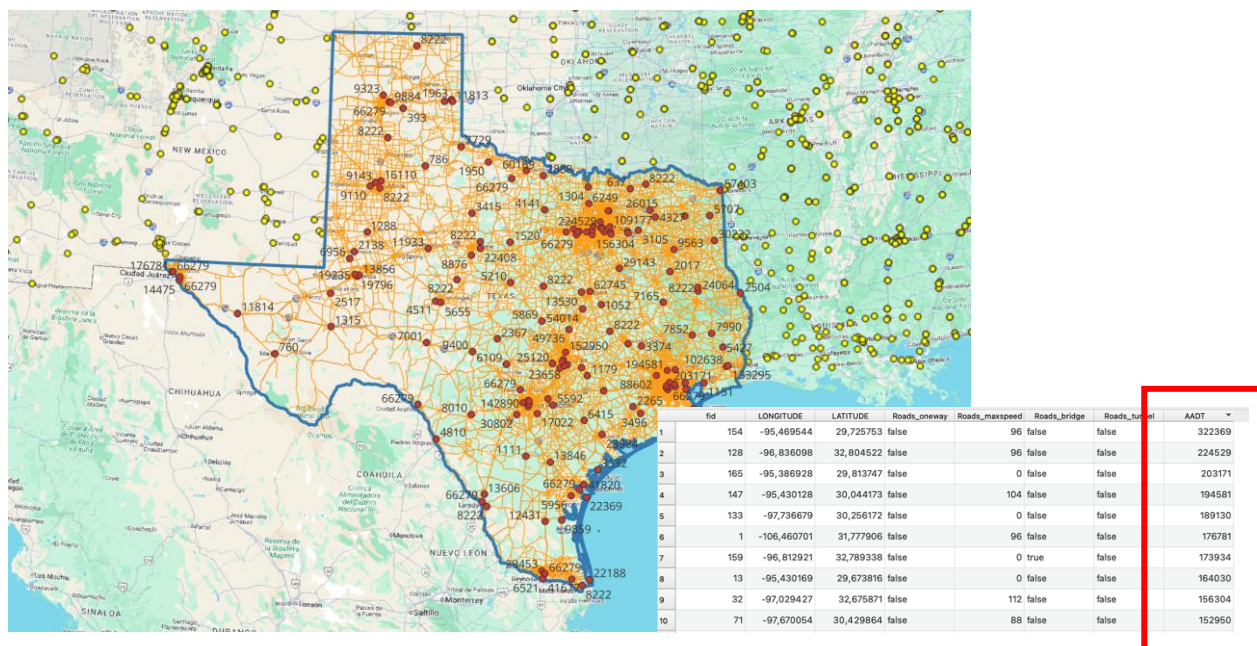
Gambar 3. Evaluasi Model

3.6 Hasil Imputasi Data

Model yang telah dilatih kemudian digunakan untuk melakukan imputasi terhadap nilai AADT yang hilang. Hasilnya, seluruh nilai AADT yang sebelumnya bernilai 0 berhasil diisi sehingga dataset menjadi lengkap (100%). Proses ini menunjukkan bahwa metode *Random Forest* mampu digunakan untuk mengatasi permasalahan *missing values* pada data lalu lintas.

3.7 Visualisasi dan Validasi Spasial

Hasil imputasi kemudian divisualisasikan dalam bentuk peta untuk melihat distribusi nilai AADT secara spasial. Berdasarkan hasil visualisasi, terlihat bahwa nilai AADT yang tinggi terkonsentrasi pada jalan utama dan wilayah metropolitan, sedangkan nilai yang lebih rendah berada pada jalan sekunder. Pola ini sesuai dengan kondisi nyata, sehingga menunjukkan bahwa hasil imputasi yang dilakukan cukup valid secara spasial.



Gambar 4. Visualisasi Hasil

3.8 Analisis Pembahasan

Berdasarkan hasil yang diperoleh, dapat diketahui bahwa pendekatan berbasis fitur geospasial mampu memberikan kontribusi dalam mengestimasi nilai AADT yang hilang. Meskipun nilai R^2 belum tinggi, model tetap mampu menghasilkan estimasi yang logis dan konsisten secara spasial. Hal ini menunjukkan bahwa atribut infrastruktur jalan memiliki peran penting dalam menentukan volume lalu lintas. Namun demikian, keterbatasan model terletak pada tidak dimasukkannya variabel lain seperti faktor temporal, kepadatan penduduk, dan aktivitas ekonomi yang juga memengaruhi lalu lintas.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengatasi permasalahan *missing values* pada data *Annual Average Daily Traffic* (AADT) melalui pendekatan *machine learning* berbasis fitur geospasial. Berdasarkan hasil identifikasi data pada wilayah Texas, ditemukan sejumlah titik pengamatan dengan nilai AADT sebesar 0 yang secara logis tidak merepresentasikan kondisi lalu lintas sebenarnya pada ruas jalan utama. Kondisi tersebut diinterpretasikan sebagai anomali atau ketidaklengkapan data akibat gangguan sensor atau proses pencatatan. Melalui integrasi data *Automatic Traffic Recorder* (ATR) dan atribut jaringan jalan dari OpenStreetMap (OSM), penelitian ini membangun fitur prediktif yang merepresentasikan karakteristik fisik dan spasial infrastruktur jalan. Model *Random Forest* yang dikembangkan mampu menjelaskan 54,8% variasi nilai AADT berdasarkan fitur tersebut. Meskipun nilai koefisien determinasi belum menunjukkan tingkat prediksi yang sangat tinggi, hasil ini mengindikasikan bahwa atribut geospasial memiliki kontribusi signifikan dalam menjelaskan distribusi volume lalu lintas. Proses imputasi yang dilakukan berhasil mengisi seluruh nilai AADT yang hilang, sehingga menghasilkan dataset dengan tingkat kelengkapan 100%. Validasi spasial melalui visualisasi peta menunjukkan bahwa hasil estimasi membentuk pola distribusi yang konsisten dengan hierarki jaringan jalan dan konsentrasi wilayah metropolitan. Dengan demikian, pendekatan berbasis fitur geospasial terbukti efektif dalam meningkatkan integritas dan konsistensi spasial data lalu lintas. Penelitian ini memberikan kontribusi dalam integrasi data sensor lalu lintas dan atribut infrastruktur jalan terbuka untuk keperluan imputasi data, serta membuka peluang pengembangan model yang lebih komprehensif dengan memasukkan variabel temporal dan sosio-ekonomi pada penelitian selanjutnya.

REFERENCES

- [1] A. Y. Nugroho, "Analisis Tingkat Kemacetan Lalu Lintas Menggunakan Data CCTV Pemerintah Kabupaten Sleman Berbasis Sistem Informasi," *JIMU: Jurnal Ilmiah Multidisipliner*, vol. 04, no. 01, 2025, [Online]. Available: <https://24jam.slemankab.go.id/>
- [2] J. Jusnidar, R. Riskayanti, Y. Setiawan, K. Khomsiah, A. Diath Yogmar, and D. Tribuana, "IMPLEMENTASI WEB-BASED TRANSPORTATION MANAGEMENT SYSTEM DI INDUSTRI TRANSPORTASI: SEBUAH TINJAUAN SISTEMATIS LITERATUR," *Jurnal Teknologi dan Bisnis Cerdas*, vol. 1, no. 3, pp. 194–209, Dec. 2025, doi: 10.64476/jtbc.v1i3.15.
- [3] Abdullah, L. P. Wardani, and Y. Siregar, "Prediksi Kemacetan Kota Semarang Berbasis Artificial Intelligence Sebagai Dasar Kebijakan Transportasi Cerdas Dan Smart Mobility," *Jurnal Mahasiswa*, vol. 7, no. 4, 2025, doi: 10.51903/jurnalmahasiswa.v7i4.1438.
- [4] P. E. Syaputra, "KAJIAN INTEGRASI TRANSPORTASI MULTI MODA UNTUK MENEKAN BIAYA LOGISTIK PADA WILAYAH KEPULAUAN (STUDI KASUS PADA PULAU BAWEAN)," *Jurnal Transportasi*, vol. 24, no. 1, pp. 49–61, 2024.
- [5] D. Budi Hartanto, "PREDIKSI KUALITAS UDARA MENGGUNAKAN MODEL KOMBINASI ALGORITMA GAUSSIAN PROCESS REGRESSION DAN GENETIC ALGORITHM," *JURNAL INFORMATIKA TEKNOLOGI DAN SAINS (JINTEKS)*, vol. 7, no. 4, pp. 1660–1668, 2025.
- [6] V. Anggraini, M. I. Faturohman, and H. Q. Karima, "Perencanaan Rute Distribusi Logistik Perkotaan Berbasis Data Geospasial Menggunakan Pendekatan Vehicle Routing Problem," *WALUYO JATMIKO PROCEEDING*, vol. 18, no. 1, pp. 451–468, Jan. 2026, doi: 10.33005/wj.v18i1.186.
- [7] A. Fikri, "Prediksi Kemacetan Lalu Lintas Urban Menggunakan Model Pembelajaran Mesin dan Data Mobilitas Real-time," *Journal of Engineering and Technological Science*, vol. 1, no. 2, pp. 61–67, 2025, doi: 10.70716/jets.v1i2.133.
- [8] A. H. Nurmakhluhi and E. Zuliarso, "PENERAPAN MACHINE LEARNING DALAM ANALISIS POLA PENGGUNAAN BANDWIDTH PADA PROVIDER INTERNET," *JURNAL INOVTEK POLBENG - SERI INFORMATIKA*, vol. 10, no. 3, pp. 1331–1341, 2025.
- [9] M. J. Faisti, R. Hadapiningradja Kusumodestoni, G. Wahyu, and N. Wibowo, "Mental Health Classification Using Naïve Bayes and Random Forest Algorithms," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 4, p. 1740, 2025, [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [10] R. Arisandi, "PERBANDINGAN MODEL KLASIFIKASI RANDOM FOREST DENGAN RESAMPLING DAN TANPA RESAMPLING PADA PASIEN PENDERITA GAGAL JANTUNG," *Jurnal Gaussian*, vol. 12, no. 1, pp. 136–145, May 2023, doi: 10.14710/j.gauss.12.1.136-145.
- [11] B. Riznov and J. Santoso, "Pembangunan Model Klasifikasi untuk Menilai Lokasi Strategis Usaha Laundry Berdasarkan Analisis Geospasial dan Data Open-Source," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 6, no. 1, pp. 73–84, 2026, doi: 10.57152/malcom.v6i1.2358.
- [12] H. Utama, A. Masruro, and T. Indriyatmoko, "Integrasi Augmentasi Data dan Machine Learning dalam Prediksi Magnitudo Gempa Bumi," *BHATARA: Jurnal Multidisiplin*, vol. 2, no. 3, pp. 97–108, 2025, [Online]. Available: <https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php>
- [13] W. Setyadi, Y. Nurhadryani, and I. Hermadi, "Pengembangan Sistem Manajemen Data Spasial Aset Jalan Tol (Studi Kasus Ruas Jalan Tol Bakauheni-Terbanggi Besar)," *JGISE: Journal of Geospatial Information Science and Engineering*, vol. 7, no. 1, p. 62, Aug. 2024, doi: 10.22146/jgise.97651.
- [14] M. Gopinda, "STUDI PERBANDINGAN ALGORITMA MACHINE LEARNING UNTUK PREDIKSI PENJUALAN RETAIL," *Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 1, no. 1, pp. 22–28, 2026, [Online]. Available: <https://ojs.feliciajurnal.com/index.php/jikti/index>
- [15] D. Manurung, B. Zealtiel, and A. H. Lubis, "Prediksi Produksi Tanaman Padi di Indonesia dengan Menggunakan Algoritma Random Forest Regressor," *Journal of Computing and Informatics Research*, vol. 4, no. 3, p. 345, 2025, doi: 10.47065/comforch.v4i3.2125.
- [16] Z. Arsyad and S. Umaroh, "Penerapan Geo-Spatial Business Intelligence dalam Analisis dan Visualisasi Survey Kondisi Jalan," *INTERNAL (Information System Journal)*, vol. 7, no. 2, pp. 152–166, 2024, doi: 10.32627.
- [17] M. R. Rahman, B. Jaya, and R. Latief, "Pengaruh Nilai Tanah Terhadap Rencana Detail Tata Ruang Di Kawasan Emas Garongkong Kabupaten Barru," *Urban and Regional Studies Journal*, vol. 8, no. 1, pp. 94–109, Dec. 2025, doi: 10.35965/ursj.v8i1.7645.
- [18] A. Riansah, O. Nurdiawan, and R. Herdiana, "PENERAPAN ALGORITMA RANDOM FOREST DAN DECISION TREE UNTUK MENINGKATKAN AKURASI KLASIFIKASI PENJUALAN PADA TOKO BANGUNAN," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 4242–4249, 2025.
- [19] A. P. Argadianata, D. A. Fatah, and H. Sukri, "KLASIFIKASI KUALITAS BUAH APEL MENGGUNAKAN METODE RANDOM FOREST," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, pp. 2016–2022, 2025.
- [20] E. R. Br Sebayang, Y. H. Chrisnanto, and Melina, "Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest," *IJESPG Journal*, vol. 1, no. 3, pp. 237–253, 2023, [Online]. Available: <http://ijespgjournal.org>
- [21] S. S. Putro *et al.*, "Classification of diabetes mellitus disease at Rato Ebu Hospital-Indonesia using the K-Nearest neighbors method based on missing value," *BIO Web Conf.*, vol. 146, Nov. 2024, doi: 10.1051/bioconf/202414601081.