TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1071-1077
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8806

Prediksi Kemacetan Lalu Lintas Menggunakan Algoritma Random Forest

Raditia Vindua, Salsa Sayida Bilqis, Hamdan Qo'du Ilal Hakim, Rido Ramadhani*

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
Email: [1]dosen02380@unpam.ac.id, [2]salsabilqis19@gmail.com, [3]hamdanqodu@gmail.com, [4,*]ridoramadhani29@gmail.com
Email Penulis Korespondensi: ridoramadhani29@gmail.com

Abstrak–Kemacetan lalu lintas merupakan permasalahan utama di wilayah urban dan suburban seperti Pamulang, Tangerang Selatan, yang disebabkan oleh peningkatan volume kendaraan tanpa diimbangi pengembangan infrastruktur jalan. Penelitian ini bertujuan untuk memprediksi tingkat kemacetan lalu lintas menggunakan algoritma Random Forest serta mengidentifikasi faktor dominan yang memengaruhi kemacetan. Data penelitian diperoleh dari Smart Traffic Management Dataset yang berisi 2.000 observasi dengan 12 variabel seperti kecepatan rata-rata, volume kendaraan, kondisi cuaca, dan laporan kecelakaan. Proses analisis mencakup tahap pembersihan data, encoding variabel kategorikal, pembagian data (80% pelatihan dan 20% pengujian), serta validasi menggunakan teknik 5-fold cross-validation. Hasil penelitian menunjukkan bahwa model Random Forest mampu memprediksi tingkat kemacetan dengan akurasi 95,75%, precision 0.96, dan recall 0.96. Analisis feature importance mengindikasikan bahwa variabel kecepatan rata-rata kendaraan memiliki kontribusi terbesar terhadap tingkat kemacetan, diikuti oleh volume kendaraan dan laporan kecelakaan. Temuan ini menunjukkan bahwa algoritma Random Forest efektif dalam mendeteksi pola lalu lintas dan dapat digunakan sebagai dasar pengembangan sistem transportasi cerdas di wilayah Pamulang untuk mengurangi kemacetan secara adaptif dan real-time.

Kata Kunci: Random Forest; Kemacetan Lalu Lintas; Pamulang

Abstract–Traffic congestion is a major issue in urban and suburban areas such as Pamulang, South Tangerang, caused by increasing vehicle volume without proportional road infrastructure development. This study aims to predict traffic congestion levels using the Random Forest algorithm and identify the dominant factors influencing congestion. The research utilized the Smart Traffic Management Dataset containing 2,000 observations with 12 variables such as average speed, traffic volume, weather conditions, and accident reports. The analysis process included data cleaning, categorical variable encoding, data splitting (80% training and 20% testing), and model validation using the 5-fold cross-validation technique. The results showed that the Random Forest model achieved 95,75% accuracy, with a precision of 0.96 and recall of 0.96. The feature importance analysis indicated that the average vehicle speed variable had the highest influence on congestion levels, followed by traffic volume and accident reports. These findings demonstrate that the Random Forest algorithm is highly effective in identifying traffic patterns and can serve as a foundation for developing intelligent transportation systems in the Pamulang area to reduce congestion adaptively and in real time.

Keywords: Random Forest; Traffic Congestion; Traffic Prediction; Pamulang

1. PENDAHULUAN

Pertumbuhan populasi dan peningkatan jumlah kendaraan pribadi telah menjadi pemicu utama permasalahan transportasi di berbagai kota besar di dunia, tidak terkecuali di Indonesia. Wilayah penyangga Jakarta, seperti Pamulang di Tangerang Selatan, mengalami perkembangan pesat yang berdampak langsung pada volume lalu lintas harian. Peningkatan volume kendaraan yang tidak diimbangi dengan pengembangan infrastruktur jalan yang memadai menyebabkan kemacetan menjadi fenomena yang tidak terhindarkan. Kemacetan lalu lintas tidak hanya menyebabkan pemborosan waktu dan bahan bakar, tetapi juga meningkatkan tingkat stres pengendara serta polusi udara yang berdampak buruk bagi lingkungan dan kesehatan masyarakat (Niron et al., n.d.). Lebih jauh, permasalahan kemacetan dapat memengaruhi produktivitas ekonomi masyarakat, karena waktu tempuh perjalanan menjadi semakin panjang. Oleh karena itu, diperlukan solusi cerdas yang mampu mengelola dan memprediksi kondisi lalu lintas secara efektif sebagai upaya preventif dan mitigatif.

Perkembangan teknologi informasi dan kecerdasan buatan (*Artificial Intelligence*) menawarkan pendekatan baru untuk mengatasi masalah kemacetan. Salah satu implementasinya adalah melalui Sistem Transportasi Cerdas *(Intelligent Transportation System – ITS*), yang bertujuan untuk mengoptimalkan kinerja sistem transportasi modern. Komponen kunci dari ITS adalah kemampuan untuk memprediksi kondisi lalu lintas secara akurat dan real-time berdasarkan data historis maupun data sensor yang terintegrasi (Farhan Raihan Wahidin et al., 2025) . Dengan adanya prediksi yang tepat, pihak berwenang dapat mengambil tindakan preventif seperti rekayasa lalu lintas, pengaturan waktu lampu lalu lintas, serta penyediaan informasi kepada pengguna jalan. Di sisi lain, pengendara juga dapat memilih rute alternatif untuk menghindari titik kemacetan sehingga mampu menekan penumpukan kendaraan pada jam-jam sibuk.

Sejumlah penelitian sebelumnya telah mengeksplorasi penggunaan berbagai algoritma *machine learning* untuk prediksi kemacetan. Sebagai contoh, penelitian (Febrianto et al., 2024) menerapkan algoritma K-Nearest Neighbors (K-NN) untuk memprediksi situasi lalu lintas dan berhasil mencapai akurasi 85%. Studi lain oleh (Yao & Ye, 2020) menggunakan Naïve Bayes dan model linier untuk klasifikasi pembagian arus lalu lintas berdasarkan waktu perjalanan. Upaya perbandingan juga dilakukan oleh (Rahayu et al., 2023)dengan menguji metode K-NN, Naïve Bayes, dan Decision Tree untuk mengklasifikasikan tingkat kemacetan di Jakarta, di mana Decision Tree menunjukkan kinerja terbaik dalam interpretabilitas model. Adapun studi komparasi lainnya oleh (Pranadjaya et al., 2024) yang membandingkan *Logistic Regression, Neural Network, dan Support Vector Machine (SVM)* untuk klasifikasi jenis kendaraan, menunjukkan bahwa setiap algoritma memiliki keunggulan masing-masing tergantung pada karakteristik dataset.

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1071-1077
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8806

Meskipun metode-metode tersebut berhasil, studi-studi terkini menunjukkan bahwa metode *ensemble learning*, seperti *Random Forest*, sering kali memiliki keunggulan dalam menangani dataset yang kompleks dan non-linear. *Random Forest* adalah algoritma yang membangun sejumlah besar pohon keputusan (*decision tree*) pada saat pelatihan dan menghasilkan prediksi berdasarkan modus dari masing-masing pohon. Keunggulan utama algoritma ini adalah ketahanan terhadap *overfitting* dibandingkan dengan pohon keputusan tunggal, serta kemampuan untuk mengukur tingkat kepentingan atau kontribusi dari setiap variabel prediktor melalui *feature importance* (Dima et al., 2025). Kemampuan *feature importance* ini sangat krusial, seperti yang ditunjukkan oleh (Febrianto et al., 2024) yang menggunakan *Random Forest* untuk menganalisis pemilihan transportasi, di mana *feature importance* berhasil mengidentifikasi faktor-faktor utama dalam preferensi pengguna. Selain itu, penelitian oleh (Kiswanto et al., 2025) juga membuktikan bahwa *Random Forest* mampu meningkatkan akurasi model prediksi lalu lintas hingga 7% dibanding *Decision Tree.*

Keunggulan akurasi *Random Forest* juga telah dibuktikan dalam berbagai domain. Sebuah studi oleh (Rodji et al., 2025) membandingkan empat algoritma untuk klasifikasi keparahan kecelakaan, menunjukkan bahwa *Random Forest* memberikan akurasi tertinggi (84%) dibandingkan *Naïve Bayes*, *Decision Tree*, dan SVM. Penelitian (Mulyani & Arifin, 2025) mengonfirmasi bahwa *Random Forest* yang dikombinasikan dengan hyperparameter tuning mampu mengungguli *Decision Tree* dalam mendeteksi penyakit stroke. Bahkan ketika dihadapkan dengan algoritma ensemble kuat lainnya seperti XGBoost, *Random Forest* menunjukkan performa yang sangat kompetitif dan terkadang lebih unggul dalam beberapa metrik evaluasi (Gullam Almuzadid & Egia Rosi Subhiyakto, 2025). Hasil serupa juga ditemukan oleh (Zul et al., 2025) yang menyebutkan bahwa stabilitas *Random Forest* terhadap data noisy menjadi alasan tingginya adopsi algoritma ini. Meskipun banyak penelitian telah dilakukan, terdapat celah (*research gap*) dalam penerapan model yang spesifik untuk karakteristik lalu lintas di wilayah suburban seperti Pamulang. Selain itu, banyak penelitian berfokus pada akurasi tanpa menganalisis faktor-faktor pemicu kemacetan secara mendalam, padahal hal tersebut merupakan salah satu kekuatan dari *Random Forest.* Untuk memaksimalkan potensinya, teknik seperti hyperparameter tuning dan *feature engineering* juga penting untuk dipertimbangkan, sebagaimana dibuktikan oleh (Pranata et al., 2024)yang menunjukkan peningkatan kinerja model melalui penerapan tuning serta pemilihan fitur yang tepat. Penelitian ini mencoba mengisi celah tersebut. *State of the art* dari penelitian ini terletak pada penerapan *Random Forest* tidak hanya untuk mencapai akurasi prediksi yang tinggi, tetapi juga untuk memberikan interpretasi mendalam terhadap faktor penyebab kemacetan melalui analisis *feature importance* pada dataset lalu lintas yang komprehensif.

Tujuan utama dari penelitian ini adalah merancang dan mengevaluasi model *Random Forest* untuk memprediksi tingkat kemacetan di wilayah Pamulang. Kontribusi penelitian ini ada tiga, di antaranya menghasilkan model prediksi dengan akurasi tinggi yang dapat diandalkan, memberikan pemahaman mendalam tentang faktor-faktor dominan yang memengaruhi kemacetan di lokasi studi melalui analisis *feature importance*, serta menyajikan hasil penelitian yang dapat menjadi dasar bagi pemerintah daerah dan pengembang aplikasi dalam membangun sistem pemantauan lalu lintas yang lebih proaktif, adaptif, dan real-time guna mengurangi dampak kemacetan di masa depan.

2. METODOLOGI PENELITIAN

2.1 Lokasi dan Objek Penelitian

Penelitian ini dilakukan pada wilayah Pamulang, Tangerang Selatan, yang merupakan kawasan urban dengan tingkat mobilitas tinggi dan potensi kemacetan signifikan. Fokus penelitian diarahkan pada proses prediksi tingkat kemacetan berdasarkan variabel-variabel lalu lintas yang tersedia dalam dataset *Smart Traffic Management*.

2.2 Tahapan Penelitian

Tahapan penelitian ini dimulai dengan proses pengumpulan data dari *Smart Traffic Management Dataset* yang berisi informasi lalu lintas terkait kondisi jalan di wilayah urban. Data yang diperoleh kemudian melalui tahap pra-pemrosesan untuk memastikan kualitas dan kesiapan data sebelum digunakan dalam model. Selanjutnya, dataset dibagi menjadi data pelatihan dan data pengujian untuk membangun serta mengevaluasi kinerja model. Pada tahapan berikutnya, model *Random Forest* dilatih menggunakan data pelatihan dengan parameter yang telah ditentukan sebelumnya. Terakhir, dilakukan evaluasi model untuk menilai akurasi dan efektivitas model dalam menghasilkan prediksi yang reliebel. Seluruh tahapan ini disusun secara sistematis untuk memastikan proses penelitian berjalan terstruktur dan menghasilkan model yang dapat digunakan sebagai dasar analisis kemacetan lalu lintas.

2.3 Pengumpulan Data

Data penelitian diperoleh *dari Smart Traffic Management Dataset* yang tersedia di platform Kaggle, dengan total 2.000 baris dan 12 variabel. Dataset berisi informasi terkait kondisi lalu lintas seperti volume kendaraan, kecepatan rata-rata, cuaca, kelembapan, suhu, dan riwayat kecelakaan.Penggunaan data sekunder diperbolehkan dalam penelitian prediktif karena efisien dan sesuai untuk model pembelajaran mesin (Kusuma et al., 2025)

2.4 Pra-pemrosesan Data

Tahapan pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum digunakan dalam pelatihan model. Preoses ini mencakup pemeriksaan duplikasi, pengecekan konsistensi tipe data, serta transformasi variabel kategorial menjadi

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1071-1077
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8806

numerik menggunakan teknik *one-hot encoding*. Selain itu, dilakukan pembuatan variabel target bernama *congestion_level* yang dibentuk berdasarkan kecepatan rata-rata kendaraan dan volume lalu lintas. Nilai kecepatan rendah dan volume tinggi dikategorikan sebagai "macet", sedangkan kondisi lainnya dikategorikan sebagai "tidak macet". Seluruh fitur numerik dinormalisasi menggunakan StandarScaler untuk menyamakan rentan nilai antar fitur input. Tahapan ini oenting untuk memastikan kualitas prediksi model dan meningkatkan stabilitas proses pelatihan.

2.5 Pembagian Data

Data dibagi menjadi dua bagian: 80% sebagai data pelatihan *(training set)* dan 20% sebagai data pengujian *(testing set)*. Pembagian ini dilakukan secara acak menggunakan fungsi *train_test_split* untuk menjaga keseimbangan kelas target. Dataset dibagi menjadi 80% data latih dan 20% data uji dengan metode *train_test_split*. Pembagian ini umum digunakan agar model dapat belajar secara optimal dan diuji secara objektif (Kurniawan & Salam, 2025).

2.6 Pelatihan Model Random Forest

Algoritma *Random Forest* digunakan sebagai motode utama untuk memprediksi tingkat kemacetan. *Random Forest* dipilih karena memiliki kemampuan dalam menangani data beragam, tahan terhadap *noise*, serta mampu mempelajari hubungan kompleksantar variabel. Model dilatih menggunakan parameter *n_estimators* = 100, *max_depth* yang dibatasi untuk menghindari *overfitting*, serta *class_weight = 'balanced'* untuk mengatasi ketidakseimbangan kelas pada dataset. Dengan pendekatan ini, setiap kelas memperoleh bobot penilaian yang proporsional selama proses pelatihan sehingga menghasilkan performa prediksi yang lebih akurat.
Persamaan dasar *Random Forest* ditunjukkan pada Persamaan (1)

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_n(x) \quad ()$$

di mana $h_i(x)$ adalah hasil prediksi dari pohon keputusan ke-i, dan hasil akhir dittentukan berdasarkan mekanisme pemungutan suara mayoritas (*majority voting*). Setelah model selesai dilatih, dilakukan evaluasi performa menggunakan data pengujian untuk mengukur akurasi, *precision*, *recall*, serta *f1-score*.

2.7 Evaluasi Model

Evaluasi model dilakukan menggunakan *classification report* dan *confusion matrix* untuk mengukur performa model dalam mengklasifikasikan dua kelas target: tidak macet dan macet. Akurasi digunakan untuk melihat performa keseluruhan model, sedangkan iprecision, recall, dan *fi-score* digunakan untuk melihat sensitivitas model dalam mendeteksi kelas minoritas. Analisis *features importance* juga dilakukan untuk mengidentifikasi variable yang paling berpengaruh terhadap prediksi-prediksi kemacetan. Proses evaluasi ini bertujuan memastikan bahwa model tidak hanya memiliki akurasi tinggi, tetapi juga mampu menggeneralisasi pola data dengan baik.

Tabel 1. Variabel Faktor Kemacetan

No	Nama	Nomor	Failed	Dtype
1	*timestamp*	V1	2.000	object
2	*location_id*	V2	2.000	int64
3	*traffic_volume*	V3	2.000	int64
4	*avg_vehicle_speed*	V4	2.000	float64
5	*vehicle_count_cars*	V5	2.000	int64
6	*vehicle_count_trucks*	V6	2.000	int64
7	*vehicle_count_bikes*	V7	2.000	int64
8	*weather_condition*	V8	2.000	object
9	*temperature*	V9	2.000	float64
10	*humidity*	V10	2.000	float64
11	*accident_reported*	V11	2.000	int64
12	*signal_status*	V12	2.000	object

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 2.000 data pengamatan dengan variabel utama seperti *traffic_volume*, *avg_vehicle_speed*, *vehicle_count_cars*, *vehicle_count_trucks*, *vehicle_count_bikes*, *weather_condition*, *temperature*, *humidity*, *accident_reported*, dan *signal_status*. Pemeriksaan *missing values* menunjukkan bahwa seluruh variabel memiliki data lengkap, sehingga keseluruhan dataset dapat langsung digunakan dalam proses pelatihan model tanpa memerlukan tahapan imputasi.

Untuk memberikan gambaran lebih jelas mengenai karakteristik data, Tabel 2 menyajikan ringkasan statistik dari variabel numerik utama yang digunakan dalam penelitian.

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1071-1077
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8806

Tabel 2. Ringkasan Statistik Variabel Utama Dataset

No	Variabel	Min	Maks	Rata-rata
1	*traffic_volume*	50	998	540.96
2	*avg_vehicle_speed*	20.01	79.97	50.01
3	*vehicle_count_cars*	20	899	449.87
4	*vehicle_count_trucks*	0	99	49.27
5	*vehicle_count_bikes*	0	49	24.12
6	*temperature*	10.03	34.99	22.45
7	*humidity*	40	95	68.20

Tabel 2 memberikan gambaran umum mengenai distribusi nilai pada variabel numerik yang berperan penting dalam proses prediksi kemacetan.

3.2 Hasil Pelatihan Model

Model Random Forest dilatih menggunakan 80% data dan diuji menggunakan 20% data. Untuk mengatasi ketidakseimbangan kelas, digunakan parameter *class_weight='balanced'*, sehingga kontribusi kelas minoritas mendapat bobot yang proporsional selama proses pelatihan. Berdasarkan hasil pengujian, model menunjukkan performa yang sangat baik dengan akurasi sebesar 0.9575. Nilai evaluasi lengkap disajikan pada Tabel 3.

Tabel 3. Hasil Evaluasi Model *Random Forest*

Metrik Evaluasi	Nilai
Akurasi	0.9575
Precision kelas 0	0.99
Precision kelas 1	0.77
Recall kelas 0	0.96
Recall kelas 1	0.96
F1-Score kelas 0	0.98
F1-Score kelas 1	0.85

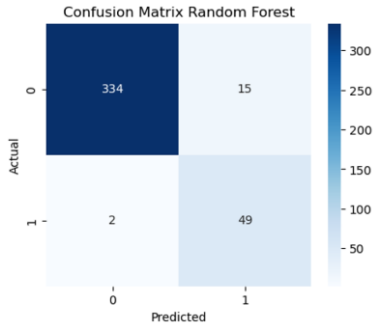
Hasil tersebut menunjukkan bahwa model mampu mengenali kedua kelas dengan baik, termasuk kelas minoritas (macet). Secara umum, model memiliki tingkat kesalahan prediksi yang rendah dan performa yang konsisten pada beberapa metrik evaluasi.

Sebagai tambahan, proses pembangunan model mengikuti beberapa tahapan utama, yaitu: (1) pra-pemrosesan data dan pembentukan variabel target, (2) pembagian data ke dalam data latih dan data uji dengan *stratified sampling*, (3) penyeimbangan kelas menggunakan *class_weight='balanced'*, (4) pelatihan model melalui mekanisme *bagging* dan *majority voting* pada sejumlah pohon keputusan, serta (5) evaluasi performa menggunakan akurasi, *precision*, *recall*, *f1-score*, dan confusion matrix. Ringkasan tahapan ini memberikan gambaran alur penyelesaian algoritma tanpa mengulangi detail teknis yang telah dibahas pada bagian Metode Penelitian.

3.3 Analisis Confusion Matrix

Hasil confusion matrix yang ditampilkan pada Gambar 1 menunjukkan bahwa model mampu memberikan prediksi yang sangat baik pada kedua kelas. Sebanyak 96% data pada kelas macet dan 96% data pada kelas tidak macet berhasil diklasifikasikan dengan benar. Keseimbangan ini menunjukkan bahwa model tidak hanya akurat dalam mengenali kelas mayoritas, tetapi juga cukup sensitif terhadap kelas minoritas.

Pencapaian ini menegaskan bahwa penerapan *class balancing* melalui penggunaan *class_weight='balanced'* berperan penting dalam meningkatkan kemampuan model dalam mendeteksi kemacetan. Selain itu, sedikitnya jumlah kesalahan prediksi menunjukkan bahwa model dapat mempelajari karakteristik tiap kelas dengan baik. Dengan demikian, confusion matrix memberikan gambaran yang lebih komprehensif mengenai kualitas klasifikasi yang dicapai oleh model.



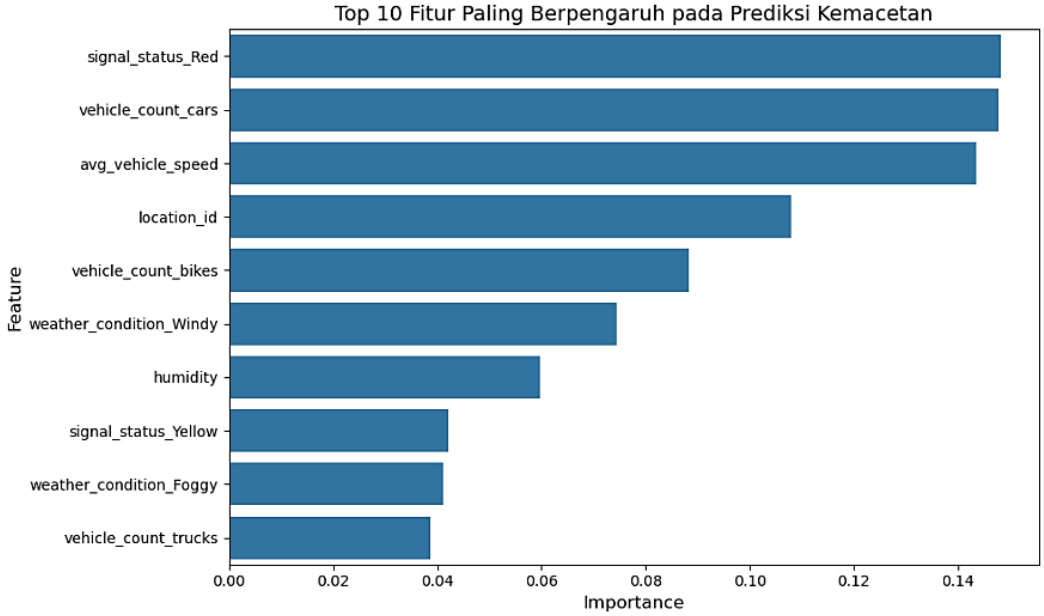
Gambar 1. *Confusion Matrix* Hasil Prediksi

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1071-1077
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8806

3.4 Analisis Feature Importance

Hasil analisis *feature importance* menunjukkan bahwa *avg_vehicle_speed* merupakan variabel yang paling berpengaruh dalam memprediksi kemacetan. Variabel ini penting karena penurunan kecepatan kendaraan secara langsung mencerminkan meningkatnya kepadatan lalu lintas. Faktor berikutnya adalah *traffic_volume*, yang menggambarkan jumlah kendaraan pada suatu waktu dan berperan besar dalam menentukan kondisi jalan.

Selain itu, jumlah kendaraan, khususnya mobil dan truk, serta variabel *accident_reported* dan *signal_status* turut memberikan kontribusi penting. Kejadian kecelakaan dan pengaturan sinyal lalu lintas umumnya memengaruhi kelancaran arus kendaraan, sehingga relevan dalam proses prediksi. Secara keseluruhan, fitur-fitur tersebut sesuai dengan pola kemacetan di lapangan yang kerap dipengaruhi oleh interaksi antara kecepatan, volume kendaraan, dan gangguan situasional.



Gambar 2. *Feature Importance Variabel*

3.5 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma Random Forest mampu menghasilkan performa prediksi yang sangat baik dalam mengidentifikasi kondisi kemacetan di wilayah Pamulang. Akurasi pengujian sebesar 0.9575, disertai nilai *precision*, *recall*, dan *f1-score* rata-rata sebesar 0.96, mengindikasikan bahwa model dapat membedakan kondisi macet dan tidak macet dengan tingkat ketepatan yang tinggi. Performa ini menunjukkan bahwa Random Forest cocok untuk permasalahan klasifikasi lalu lintas yang memiliki pola data kompleks.

Kualitas generalisasi model juga terlihat dari hubungan antara akurasi pada data pelatihan sebesar 0.8731 dan nilai rata-rata *cross-validation* yang mencapai 0.9455. Kedua nilai ini tidak berbeda secara signifikan, sehingga dapat disimpulkan bahwa model tidak mengalami *overfitting*. Konsistensi performa di berbagai pembagian data menunjukkan bahwa model dapat bekerja dengan stabil dan tidak hanya menghafal pola pada data latih.

Peningkatan akurasi model terutama terjadi setelah diterapkannya teknik penyeimbangan kelas menggunakan *class_weight='balanced'* dan *stratified sampling*. Sebelum dilakukan balancing, model cenderung sulit mengenali kelas macet yang jumlahnya jauh lebih sedikit. Setelah proses penyeimbangan dilakukan, nilai *recall* kelas macet meningkat hingga 0.96, memperlihatkan bahwa model semakin peka terhadap kondisi kemacetan dan tidak lagi bias pada kelas dengan jumlah data lebih besar.

Analisis terhadap *feature importance* mengungkapkan bahwa *avg_vehicle_speed* merupakan fitur yang paling berpengaruh, diikuti oleh *traffic_volume*, jumlah kendaraan berdasarkan jenis, serta variabel *accident_reported* dan *signal_status*. Kontribusi fitur-fitur tersebut sejalan dengan kondisi lalu lintas di Pamulang, di mana penurunan kecepatan dan peningkatan volume kendaraan sering kali menjadi pemicu terbentuknya kemacetan pada titik-titik tertentu seperti kawasan Jalan Siliwangi, Jalan Pajajaran, dan sekitar Pamulang Square. Hal ini menunjukkan bahwa model mampu merepresentasikan pola lalu lintas yang terjadi di lapangan.

Walaupun performa model tergolong sangat baik, beberapa kesalahan prediksi masih muncul pada kondisi lalu lintas yang berada di antara kategori padat dan macet. Variabel yang terbatas pada dataset tidak sepenuhnya menangkap faktor eksternal seperti hambatan sesaat atau kondisi jalan tertentu. Oleh karena itu, penelitian lanjutan dapat mempertimbangkan penggunaan data real-time, fitur temporal seperti jam puncak, serta penyempurnaan hyperparameter untuk meningkatkan akurasi model. Secara keseluruhan, algoritma Random Forest terbukti efektif untuk memprediksi kemacetan dan berpotensi diterapkan dalam sistem manajemen lalu lintas berbasis data.

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1071-1077
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8806

4. KESIMPULAN

Penelitian ini bertujuan membangun model prediksi kemacetan lalu lintas di wilayah Pamulang dengan memanfaatkan algoritma Random Forest. Berdasarkan hasil pengujian, model yang dibangun menunjukkan performa yang sangat baik dengan akurasi sebesar 95,75%. Nilai *precision*, *recall*, dan *f1-score* rata-rata yang mencapai 0.96 memperlihatkan bahwa model mampu membedakan kondisi macet dan tidak macet secara akurat. Kinerja tersebut juga konsisten dengan akurasi pelatihan sebesar 0.8731 serta nilai *cross-validation* rata-rata 0.9455, yang menunjukkan bahwa model stabil dan tidak mengalami *overfitting*. Analisis terhadap kontribusi fitur mengungkapkan bahwa *avg_vehicle_speed* dan *traffic_volume* merupakan variabel yang paling berpengaruh dalam membentuk prediksi kemacetan. Faktor tersebut diikuti oleh jumlah kendaraan berdasarkan jenis, keberadaan laporan kecelakaan, serta status sinyal lalu lintas. Temuan ini menunjukkan bahwa dinamika kemacetan di Pamulang sangat dipengaruhi oleh kombinasi antara kecepatan kendaraan dan tingkat kepadatan lalu lintas di ruas jalan tertentu. Secara keseluruhan, penelitian ini memperlihatkan bahwa algoritma Random Forest merupakan metode yang efektif untuk memprediksi kemacetan lalu lintas dan dapat diterapkan sebagai bagian dari sistem transportasi cerdas. Untuk penelitian selanjutnya, pemanfaatan data secara real-time serta penambahan fitur temporal seperti jam sibuk atau pola harian dapat membantu meningkatkan ketelitian model dalam menggambarkan perubahan kondisi lalu lintas secara lebih komprehensif.

REFERENCES

Dima, J., Hamzah, Moh. S., Tallo, C. G., & Fallo, D. Y. A. (2025). Tinjauan Literatur tentang Pemanfaatan Algoritma Greedy untuk Pencarian Jalur Terpendek. *Jurnal Kridatama Sains Dan Teknologi*, *7*(01), 519–528. https://doi.org/10.53863/kst.v7i01.1683

Farhan Raihan Wahidin, Wina Witanti, & Edvin Ramadhan. (2025). Pengontrolan Lampu Lalu Lintas Menggunakan Teknologi Deteksi Kendaraan YOLOV4. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *12*(5), 975–984. https://doi.org/https://doi.org/10.25126/jtiik.2025125

Febrianto, M. F., Priyatno, A., Adisty, H., Saputri, A. F., Amanullah, R., Pebrianti Krissella, T., Irzavika, N., & Matondang, N. H. (2024). Prediksi Situasi Lalu Lintas Menggunakan Machine Learning Dengan Algoritma K-Nearest Neighbors Classifier. *Informatik : Jurnal Ilmu Komputer*, *20*(1), 28–34. https://doi.org/10.52958/iftk.v20i1.7042

Gullam Almuzadid, & Egia Rosi Subhiyakto. (2025). Stroke Risk Classification Using the Ensemble Learning Method of XGBoost and Random Forest. *Journal of Applied Informatics and Computing*, *9*(3), 828–837. https://doi.org/10.30871/jaic.v9i3.9528

Kiswanto, D., Ramadhani, F., Surbakti, N. M., & Nasution, N. A. (2025). *Pengembangan dan Implementasi Sistem Deteksi Serangan DDoS Berbasis Algoritma Random Forest*. *6*(3). https://doi.org/10.47065/bit.v5i2.2203

Kurniawan, W. A., & Salam, M. K. A. (2025). Penggunaan Feature Space SMOTE Untuk Mengurangi Overfitting Akibat Imbalance Dataset. *Jurnal Transformatika*, *22*(2), 140–149. https://doi.org/10.26623/transformatika.v22i2.8305

Kusuma, M. R., Informasi, P. S., Mandiri, U., Sistem, P., Informasi, T., Darunnajah, U., Seleksi, F., & Learning, A. M. (2025). *Pendekatan Analisis Prediktif Regresi Menggunakan Metode Pembelajaran Mesin Untuk Memperkirakan Effort*. *10*(1), 23–34.

Mulyani, S., & Arifin, T. (2025). Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Pendekatan Machine Learning dengan Optimasi GridSearchCV. *Jurnal Informatika Polinema*, *11*(4), 577–586. https://doi.org/10.33795/jip.v11i4.7938

Niron, W. P., Kuswara, K. M., Paul, D., & Tamelan, G. (n.d.). Faktor-Faktor Penyebab Kemacetan Lalu Lintas Di Ruas Jalan Herman Fernandez Larantuka Kabupaten Flores Timur Factors Causing Traffic Construction On Herman Fernandez Road, East Flores District. *Jurnal Teknologi*, *19*(1), 2025. https://doi.org/https://ejurnal.undana.ac.id/index.php/jurnal_teknologi/article/view/22338

Pranadjaya, E., Pangestu, E. S., Sereati, C. O., Octaviani, S., & Darmawan, M. (2024). Perbandingan Algoritma Machine Learning menggunakan Orange Data Mining untuk Klasifikasi Jenis Kendaraan pada Sistem Tilang Digital. *Jurnal Elektro*, *17*(1), 41–47. https://doi.org/10.25170/jurnalelektro.v17i1.5429

Pranata, J., Agustian, S., Jasril, J., & Haerani, E. (2024). Penggunaan Model Bahasa indoBERT pada metode Random Forest untuk Klasifikasi Sentimen dengan Dataset Terbatas. *Building of Informatics, Technology and Science (BITS)*, *6*(3), 1668–1676. https://doi.org/10.47065/bits.v6i3.6335

Rahayu, S., Asmoro, B. R., & Rinal, E. (2023). Classification of Congestion in Jakarta Using KNN, Naïve Bayes and Decision Tree Method. *Jurnal Syntax Admiration*, *4*(7), 928–952. https://doi.org/10.46799/jsa.v4i7.654

Rodji, A. P., Studi, P., Ilmu, M., Nusa, U., Jakarta, M., Melayu, J. C., Timur, K. J., Ibukota, D. K., Sipil, S. T., Teknik, F., Krisnadwipayana, U., Gede, P., Bekasi, K., Tree, D., & Keparahan, K. (2025). *Klasifikasi Keparahan Kecelakaan Lalu Lintas Menggunakan Machine Learning*. *10*(1), 82–92.

Yao, J., & Ye, Y. (2020). The effect of image recognition traffic prediction method under deep learning and naive Bayes algorithm on freeway traffic safety. *Image and Vision Computing*, *103*, 103971. https://doi.org/10.1016/j.imavis.2020.103971

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1071-1077
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8806

Zul, Z., Al, H., & Subhiyakto, E. R. (2025). *Analisis Komparatif Akurasi Prediksi Kanker Payudara Menggunakan Algoritma Random Forest dan Logistic Regression*. 300–311. https://doi.org/10.33364/algoritma/v.22-1.2164