

Transportation Mode Choice for Shopping Trip Using Random Forest Algorithm

Pemilihan Moda Transportasi Untuk Perjalanan Belanja
Menggunakan Algoritma Random Forest

Kezia G. Najoan, Samuel Y. R. Rompis, Lucia I. R. Lefrandt

Dept. of Civil Engineering, Sam Ratulangi University Manado, Kampus Bahu St., 95115, Indonesia

e-mails : kezianajoan021@student.unsrat.ac.id, semrompis@unsrat.ac.id, lucia.lefrandt@unsrat.ac.id

Received: 18 Juni 2025; revised: 20 Juni 2025; accepted: 24 Juni 2025

Abstract — Manado City, home to over 460,000 residents and anchored by the Megamas business and shopping district, faces significant mobility challenges and complex traffic congestion. The rise of online transport services has provided residents with new options, yet private vehicles remain the dominant choice due to comfort and scheduling flexibility. This study aims to describe travel-actor characteristics in Manado, identify key factors influencing mode choice, and apply a Random Forest Classifier to predict preferences between private vehicles and online transport services. Primary data were collected from 203 respondents via online questionnaires and field interviews, covering demographics, trip characteristics, and choice drivers. Data were cleaned, encoded, and split into 80% training and 20% testing subsets. A default Random Forest model (100 trees) showed overfitting (100% train vs. 78% test). Hyperparameter tuning (GridSearchCV, 5-fold) improved test accuracy to 80.5%. Removing the least important feature (“trip distance”) further boosted test accuracy to 85.4% with average F_1 -score >85%. Feature importance ranked vehicle ownership (25.4%) and primary reason (22.3%) as top predictors. Recommendations include expanding fleet coverage in suburban areas, adjusting pricing for lower-income groups, and intensifying public education on online transport benefits to alleviate congestion. Future research could explore multimodal integrations and real-time data analytics to optimize urban mobility systems.

Key words— Artificial Intelligence, Mode Choice, Online Transportation, Private Vehicle, Random Forest

Abstrak — Kota Manado, dengan lebih dari 460.000 penduduk dan kawasan Megamas sebagai pusat bisnis dan perbelanjaan, menghadapi tantangan mobilitas tinggi dan kemacetan kompleks. Munculnya layanan transportasi online menawarkan alternatif, namun kendaraan pribadi masih mendominasi pilihan karena kenyamanan dan fleksibilitas waktu. Penelitian ini bertujuan menggambarkan karakteristik pelaku perjalanan di Manado, mengidentifikasi faktor penentu pilihan moda, dan menerapkan Random Forest Classifier untuk memprediksi preferensi antara kendaraan pribadi dan transportasi online. Data primer diperoleh dari 203 responden melalui kuesioner daring dan wawancara lapangan, mencakup demografi, karakteristik perjalanan, dan pendorong pilihan. Data dibersihkan, dikodekan, dan dibagi 80% training–20% testing. Model Random Forest default (100 pohon) menunjukkan overfitting (akurasi train 100% vs. test 78%). Hyperparameter tuning (GridSearchCV 5-fold) meningkatkan akurasi test menjadi 80,5%. Menghapus fitur terendah (“jarak perjalanan”) menaikkan akurasi test menjadi 85,4% dengan F_1 -score rata-rata >85%. Feature importance menempatkan kepemilikan kendaraan (25,4%) dan alasan utama (22,3%) sebagai prediktor utama. Rekomendasi meliputi perluasan

jangkauan armada di zona pinggiran, penyesuaian tarif untuk kelompok berpendapatan rendah, dan edukasi publik tentang keunggulan transportasi daring untuk mengurangi kemacetan. Penelitian lanjutan dapat mengeksplorasi integrasi multimoda dan analitik data real-time untuk mengoptimalkan sistem mobilitas perkotaan.

Kata kunci — Kecerdasan Buatan, Kendaraan Pribadi, Pemilihan Moda, Pohon Keputusan, Transportasi Online

I. PENDAHULUAN

Kota Manado sebagai ibukota Provinsi Sulawesi Utara yang memiliki 11 kecamatan dan 87 kelurahan, di mana 10 kecamatan terletak di daratan utama Pulau Sulawesi dan 1 kecamatan berbentuk kepulauan, terus mengalami pertumbuhan baik dari segi jumlah penduduk maupun aktivitas ekonomi. Dengan populasi yang terus meningkat yang saat ini mencapai lebih dari 460.000 jiwa [1], kebutuhan akan sistem transportasi yang efisien menjadi sangat mendesak. Pertumbuhan ini tidak hanya meningkatkan jumlah kendaraan pribadi, tetapi juga mengubah pola perjalanan masyarakat.

Masyarakat kini memiliki lebih banyak pilihan moda transportasi, termasuk layanan transportasi online yang semakin populer. Penggunaan angkutan kota mulai beralih ke taksi online disebabkan oleh kualitas pelayanan, aksesibilitas keterjangkauan, waktu, dan kenyamanan. [2]

Studi mengatakan sebagian besar warga Manado (96%) pernah menggunakan mikrolet dan taksi online, bahkan probabilitas memilih taksi online (88,6%) jauh lebih tinggi daripada mikrolet (11,4%), karena kemudahan pemesanan, tempat jemput fleksibel, pendingin udara (AC), dan pelayanan yang lebih baik. [3]

Meskipun layanan seperti ojek online dan taksi online menawarkan kemudahan, kenyamanan, serta aksesibilitas dalam perjalanan, pelaku perjalanan tetap mempertimbangkan aspek biaya, waktu, dan fleksibilitas. Di sisi lain, kendaraan pribadi masih menjadi pilihan banyak orang, terutama bagi mereka yang menginginkan kontrol penuh atas perjalanan mereka.

Data di kota besar seperti Jakarta menunjukkan bahwa kendaraan bermotor pribadi masih dominan digunakan oleh lebih dari separuh responden dalam perjalanan sehari-hari. Penelitian ini memperlihatkan bahwa pola pemilihan moda

dipengaruhi oleh karakteristik sosial-demografis dan ketersediaan fasilitas transportasi. [4]

Perubahan pola perjalanan ini sangat terlihat pada rute-rute di perkotaan yang menuju ke pusat bisnis dan perbelanjaan yaitu Kawasan Megamas Manado. Kawasan Megamas Manado atau dikenal dengan istilah Boulevard on Business (BoB) Manado, yang merupakan pusat bisnis dan perbelanjaan, memiliki akses yang strategis untuk didatangi oleh masyarakat Kota Manado yang ingin bepergian ke tempat ini. Tak hanya itu, Boulevard on Business (BoB) Manado ini juga menjadi salah satu pusat ekonomi dan hiburan terkemuka di Kota Manado. Terdapat jaringan kantor bank nasional, hotel, pusat perbelanjaan, toko ritel, sarana hiburan, fasilitas olahraga, serta berbagai pusat kuliner menjadikan tempat ini sebagai salah satu tempat tersibuk karena sering dikunjungi oleh sebagian besar masyarakat di Kota Manado.

Ramainya pengunjung di tempat ini tentunya berpengaruh pada kepadatan lalu lintas yang mengakibatkan mobilitas di rute ini menjadi lebih kompleks dan menuntut solusi transportasi yang dapat mengatasi kemacetan, memenuhi kebutuhan jarak tempuh yang bervariasi, dan tetap mempertahankan kenyamanan serta efisiensi waktu.

Studi menunjukkan bahwa selain biaya dan jarak, kemudahan akses juga menjadi faktor kunci dalam pemilihan moda transportasi di Kota Manado yang bersifat maritim.[5]

Meskipun kedua moda ini tersedia, pelaku perjalanan mempunyai pertimbangan dalam memilih moda transportasi yang paling efisien untuk memenuhi kebutuhan mereka sehari-hari. Sebagai contoh, untuk bepergian di jam-jam sibuk, beberapa pelaku perjalanan mungkin lebih memilih transportasi online untuk menghindari kesulitan parkir dan kemacetan, sementara yang lain tetap mengandalkan kendaraan pribadi demi kenyamanan privasi. Pola ini menunjukkan adanya variasi dalam preferensi sebagian masyarakat di Kota Manado terkait moda transportasi.

Secara teoritis, pemilihan moda diasumsikan berdasarkan prinsip utilitas maksimum individu, yaitu pengguna cenderung memilih moda yang memaksimalkan kepuasan seperti minimasi biaya dan waktu serta maksimasi kenyamanan.[6]

Banyak penelitian pemilihan moda transportasi menggunakan model logit multinomial (MNL) atau logit biner untuk menganalisis preferensi pengguna. Namun belakangan pendekatan berbasis pembelajaran mesin (machine learning) mulai banyak digunakan karena mampu menangkap hubungan kompleks nonlinier dalam data transportasi tanpa asumsi teoretis yang ketat. Penerapan metode machine learning seperti Random Forest, XGBoost, atau SVM dalam studi perilaku transportasi terbukti sering memberikan kinerja prediksi setara atau lebih baik dibanding metode diskrit konvensional.[7]

Dalam penelitian yang dilakukan oleh Sekhar et al. (2016), dijelaskan bahwa analisis pemilihan moda merupakan bagian integral dari proses perencanaan transportasi karena memberikan wawasan mendalam tentang preferensi moda pengguna dan dapat digunakan sebagai alat evaluasi untuk pengenalan sistem transportasi baru. Penelitian ini menunjukkan bahwa dengan menggunakan metode Random Forest (RF), analisis pemilihan moda dapat dilakukan dengan

akurasi yang lebih tinggi dibandingkan dengan model tradisional seperti Multinomial Logit (MNL). Hasilnya menunjukkan akurasi prediksi RF mencapai 98,96%, jauh lebih tinggi daripada Multinomial Logit (77,31%) pada data komuter Delhi, India. [8]

Studi oleh Zhang et al. (2023) memodelkan pilihan moda menggunakan stacking ensemble ML termasuk Random Forest dan membandingkannya dengan model logit multinomial. Hasilnya metode stacking yang mencakup Random Forest mencapai akurasi 83%, mengungguli model MNL konvensional serta berbagai model ML tunggal lainnya.[9]

Random Forest secara signifikan lebih akurat dalam memprediksi pilihan moda dibanding model MNL tradisional. Keakuratan RF disebabkan kemampuannya menangkap relasi non-linier yang kompleks antara atribut sosio-ekonomi dan pilihan moda. [10]

Selain kemampuan prediksi, Random Forest dapat mengidentifikasi variabel-variabel penting (misalnya biaya, waktu, kenyamanan) yang memengaruhi keputusan moda, informasi yang sangat berguna untuk perencanaan transportasi.

Penerapan metode Random Forest Classifier yang berbasis Artificial Intelligence (AI) menawarkan solusi inovatif untuk menganalisis data terkait preferensi moda transportasi secara lebih akurat. Metode ini memungkinkan identifikasi faktor-faktor kunci yang memengaruhi keputusan pelaku perjalanan dalam memilih moda transportasi. Penggunaan AI dalam analisis pemilihan moda transportasi dapat memberikan wawasan baru tentang perilaku pelaku perjalanan serta membantu merumuskan kebijakan publik yang lebih berbasis bukti.[11]

Zhang dan Liu mengembangkan metode Adaptive Waterwheel Particle Optimization untuk mengoptimalkan parameter Random Forest dalam pemilihan moda transportasi perkotaan, sehingga meningkatkan akurasi prediksi dan interpretabilitas model Random Forest dibandingkan konfigurasi default [12]. Sementara itu, studi yang lain menemukan bahwa Random Forest secara konsisten melampaui performa model diskrit konvensional (Nested/MNL), serta mengungkap variabel signifikan seperti waktu, jarak, kepemilikan kendaraan, dan karakteristik lingkungan binaan yang memengaruhi keputusan moda.[13]

Meskipun banyak penelitian terkait pemilihan moda transportasi telah dilakukan di berbagai kota besar, penelitian yang memfokuskan pada perbandingan antara transportasi online dan kendaraan pribadi di Kota Manado masih sangat terbatas. Tidak banyak penelitian yang secara eksplisit menggunakan Random Forest Classifier untuk menganalisis faktor-faktor yang memengaruhi pilihan moda transportasi di daerah ini. Dalam penelitian ini, Random Forest Classifier digunakan karena mampu menangani data yang beragam, baik angka maupun kategori, dengan hasil yang akurat. Metode ini lebih stabil dan dapat mencegah overfitting karena menggabungkan banyak pohon keputusan. Selain itu, Random Forest dapat menunjukkan faktor mana yang paling memengaruhi pilihan moda transportasi, yang sangat penting untuk penelitian ini. Dibandingkan metode lain seperti regresi logistik, Random Forest lebih fleksibel karena dapat bekerja

dengan baik meskipun data tidak memiliki hubungan linier, data mengandung variabel yang berkorelasi, dan data memiliki distribusi yang tidak normal.

Tujuan penelitian ini adalah menganalisis karakteristik pelaku perjalanan dan preferensi masyarakat Manado dalam memilih moda transportasi, menganalisis faktor yang memengaruhi keputusan pemilihan moda, dan menerapkan algoritma Random Forest Classifier untuk memprediksi preferensi antara kendaraan pribadi dan transportasi online pada rute menuju Kawasan Megamas di Kota Manado.

Dengan pendekatan data-driven, model Random Forest diharapkan tidak hanya meningkatkan akurasi prediksi pemilihan moda, tetapi juga memberikan insight baru tentang perilaku pelaku perjalanan yang penting untuk perumusan kebijakan transportasi publik di Manado.

II. METODE

A. Algoritma Random Forest Classifier

Random Forest adalah metode ensemble learning berbasis pohon keputusan yang banyak digunakan untuk tugas klasifikasi dan regresi. Algoritma ini dikembangkan oleh Leo Breiman pada tahun 2001 sebagai pengembangan dari ide bagging dan random subspace untuk mengurangi overfitting dan meningkatkan akurasi model pohon tunggal. Secara umum, Random Forest membangun decision trees secara independen, lalu mengambil keputusan akhir melalui majority voting untuk klasifikasi.[14]

Tahap pertama adalah membuat bootstrapped dataset yaitu mengambil data acak pada data train dengan pengembalian.

Tahap kedua, memilih subset fitur acak pada node..

$$m = \sqrt{X} \quad (1)$$

Dimana, m = jumlah fitur yang dipilih

X = jumlah total fitur

Tahap ketiga, membentuk root atau node pada pohon keputusan dengan perhitungan gini impurity.

$$\text{Gini} = 1 - (P_0^2 + P_1^2) \quad (2)$$

Dimana,

p_0 = probabilitas yang didapat dari jumlah kelas positif dibagi total kasus

p_1 = probabilitas yang didapat dari jumlah kelas negatif dibagi total kasus

Langkah ini dilakukan berulang sampai mendapat pohon keputusan.

Keputusan akhir dari Random Forest adalah majority voting.

B. Penerapan Random Forest

Pada penelitian ini, data primer dikumpulkan dari 203 responden yang melakukan perjalanan menuju Kawasan Megamas di Kota Manado melalui kuesioner tertutup berbasis informed consent. Variabel dependen adalah pilihan moda antara transportasi online (0) dan kendaraan pribadi (1), sedangkan variabel independen meliputi demografi (usia, jenis kelamin, pekerjaan, pendapatan, kepemilikan kendaraan, lokasi domisili), karakteristik perjalanan (tujuan, frekuensi, jarak, waktu tempuh, biaya), dan alasan pemilihan (kenyamanan,

parkir, keamanan, dsb.). Data dibersihkan yaitu data duplikat dihapus, missing value diimputasi minimal, kemudian variabel kategorik di-encode (label encoding untuk variabel biner/ordinal, one-hot encoding untuk variabel nominal).

Dataset dibagi menjadi training (80%) dan testing (20%) dengan stratified split untuk menjaga proporsi kelas moda. Model Random Forest Classifier awal diinisialisasi dengan parameter default ($n_estimators=100$, $criterion='gini'$, $max_features='sqrt'$, $bootstrap=True$, $random_state=42$). Apabila evaluasi data overfit, dilakukan hyperparameter tuning menggunakan GridSearchCV 5-fold pada data training, menguji parameter $n_estimators$ [100,200,300], max_depth [None,10,20,30], $min_samples_split$ [2,5,10], $min_samples_leaf$ [1,2,4], $max_features$ ['sqrt','log2',0.2,0.5], dan $class_weight='balanced'$ jika diperlukan. Model terpilih dilatih ulang pada keseluruhan data training dengan parameter optimal. Evaluasi performa akhir pada data testing menghitung akurasi, precision, recall, F1-score, dan confusion matrix.

Setelah model optimal diperoleh, feature importance diekstraksi dari atribut $feature_importances$ untuk mengidentifikasi variabel determinan utama pemilihan moda. Berdasarkan skor importance, dilakukan seleksi fitur opsional dengan menghapus fitur berkontribusi rendah, lalu melatih ulang model untuk memeriksa dampak pada performa.

Seluruh implementasi menggunakan Python 3.11.2 dengan scikit-learn v1.5.1, pandas, numpy, serta visualisasi internal menggunakan matplotlib. Pustaka scikit-learn dipilih karena menyediakan implementasi Random Forest stabil dan fitur tuning/validasi yang lengkap.[15]

III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data kuesioner dari 203 responden di Kota Manado yang melakukan perjalanan menuju Kawasan Megamas. Sebelum menerapkan analisis klasifikasi Random Forest untuk membedakan moda transportasi online dan kendaraan pribadi, langkah pertama yang dilakukan adalah mempelajari karakteristik pelaku perjalanan, pola perjalanannya, serta preferensi masyarakat dalam memilih moda transportasi.

A. Karakteristik Responden

Karakteristik pelaku perjalanan meliputi variabel seperti jenis kelamin, rentang umur, status pekerjaan, tingkat pendapatan, kepemilikan kendaraan pribadi, dan lokasi tempat tinggal yang proporsinya dapat dilihat pada Tabel I. Setiap variabel ini kemudian dianalisis untuk melihat preferensi antara moda transportasi online dan kendaraan pribadi.

Selanjutnya, pada karakteristik perjalanan membahas aspek-aspek perjalanan responden menuju Kawasan Megamas Manado, termasuk frekuensi perjalanan, tujuan utama, jarak tempuh, waktu tempuh, serta biaya perjalanan yang dikeluarkan (*lihat Tabel II*). Karakteristik ini penting untuk dianalisis karena mencerminkan pola mobilitas masyarakat di kawasan perkotaan dengan aktivitas ekonomi tinggi.

Kemudian, preferensi pemilihan moda akan dibahas bagaimana masyarakat Kota Manado menimbang berbagai faktor ketika memilih moda transportasi—apakah mereka lebih condong menggunakan layanan daring seperti Gojek atau Grab,

TABEL I. KARAKTERISTIK PELAKU PERJALANAN

| Variabel | Sub-Variabel | Persentase(%) |
|-----------------------|------------------------|---------------|
| Jenis kelamin | Laki-laki | 46,80% |
| | Perempuan | 53,20% |
| Usia | <20 tahun | 14,30% |
| | 20-29 tahun | 72,40% |
| | 30-39 tahun | 8,90% |
| | >40 tahun | 4,40% |
| Pekerjaan | Belum bekerja | 2,50% |
| | Buruh/jasa | 1,50% |
| | Honoror | 2,95% |
| | Ibu rumah tangga | 1,97% |
| | P3K | 0,98% |
| | PNS/TNI/POLRI | 4,92% |
| | Pedagang/wiraswasta | 6,89% |
| | Pegawai swasta | 14,77% |
| | Pelajar/mahasiswa | 63,54% |
| Pendapatan | <Rp2.000.000 | 63,54% |
| | Rp2.000.000-4.999.999 | 22,16% |
| | Rp5.000.000-7.499.999 | 9,35% |
| | Rp7.500.000-10.000.000 | 2,95% |
| | >Rp10.000.000 | 1,97% |
| Kepemilikan kendaraan | Belum memiliki | 18,22% |
| | Sepeda motor | 39,40% |
| | Mobil | 14,28% |
| | Memiliki keduanya | 28,07% |
| Lokasi domisili | Bunaken | 1,47% |
| | Malalayang | 34,48% |
| | Mapanget | 10,34% |
| | Paal dua | 5,91% |
| | Sario | 7,88% |
| | Singkil | 6,40% |
| | Tikala | 1,47% |
| | Tuminting | 2,46% |
| | Wanea | 26,60% |
| | Wenang | 2,95% |

TABEL II. KARAKTERISTIK PERJALANAN

| Variabel | Sub-Variabel | Persentase(%) |
|-----------------------|--------------------|---------------|
| Frekuensi perjalanan | Jarang | 13,30% |
| | 1-3x sebulan | 25,61% |
| | 1-2x seminggu | 33,00% |
| | 3-4x seminggu | 17,24% |
| | Hampir setiap hari | 10,83% |
| Tujuan perjalanan | Bekerja | 5,41% |
| | Belajar | 0,98% |
| | Belanja | 12,31% |
| | Bersantap | 4,43% |
| | Rekreasi | 76,84% |
| Jarak perjalanan (km) | <1 km | 2,46% |
| | 1-5 km | 56,65% |
| | 5,1-10 km | 28,07% |
| | 10,1-15 km | 8,37% |
| | 15,1-20 km | 2,46% |
| | >20 km | 1,97% |
| Waktu tempuh (menit) | 0-15 menit | 53,20% |
| | 16-30 menit | 37,43% |
| | 31-45 menit | 6,89% |
| | 45-60 menit | 1,47% |
| | >60 menit | 0,98% |
| Biaya perjalanan (Rp) | <Rp10.000 | 14,77% |
| | Rp10.000-20.000 | 46,79% |
| | Rp20.001-30.000 | 23,15% |
| | Rp30.001-40.000 | 6,89% |
| | Rp40.001-50.000 | 2,95% |
| | >Rp50.000 | 5,41% |

atau justru lebih memilih kendaraan pribadi. Analisis ini didasarkan pada data kuisioner yang menanyakan alasan utama

TABEL III. PREFERENSI PEMILIHAN MODA

| Variabel | Sub-Variabel | Persentase(%) |
|------------------------|--------------------------|---------------|
| Pemilihan Moda | Transportasi Online | 37,43% |
| | Kendaraan Pribadi | 62,56% |
| Alasan utama pemilihan | Biaya lebih terjangkau | 13,79% |
| | Lebih aman | 3,44% |
| | Mudah diakses | 17,73% |
| | Lebih nyaman | 12,80% |
| | Moda tersedia | 11,82% |
| | Tidak perlu cari parkir | 4,43% |
| | Waktu tempuh lebih cepat | 35,96% |

TABEL IV. AKURASI MODEL RANDOM FOREST DEFAULT

| Data | Akurasi |
|----------|---------|
| Training | 100% |
| Testing | 78% |

TABEL V. HASIL CLASSIFICATION REPORT

| | Precision | Recall | F1-Score |
|---|-----------|--------|----------|
| 0 | 0,71 | 0,83 | 0,77 |
| 1 | 0,85 | 0,74 | 0,79 |

TABEL VI. AKURASI MODEL SETELAH HYPERPARAMETER TUNING

| Data | Akurasi |
|----------|---------|
| Training | 84,56% |
| Testing | 80,5% |

TABEL VII. CLASSIFICATION REPORT SETELAH HYPERPARAMETER TUNING

| | Precision | Recall | F1-Score |
|---|-----------|--------|----------|
| 0 | 0,86 | 0,67 | 0,75 |
| 1 | 0,78 | 0,91 | 0,84 |

responden, mulai dari waktu tempuh, biaya perjalanan, kemudahan akses, kenyamanan, hingga keamanan (Tabel III).

B. Performa Model

Model Random Forest default menunjukkan akurasi training 100% vs. testing 78%, menandakan overfitting awal (*lihat Tabel IV*).

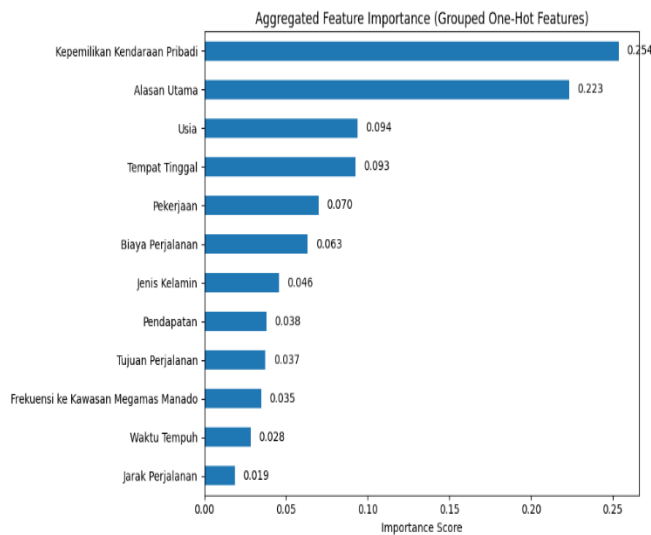
Tabel V menjelaskan bahwa dari semua prediksi 0 (online) 71% diprediksi benar artinya model cukup presisi, dari semua data aktual online 83% diprediksi benar artinya model mengenali data dengan baik (recall), rata-rata dari precision dan recall kelas online yaitu 77%. Sedangkan dari semua prediksi 1 (pribadi) 85% diprediksi benar, dari semua data aktual pribadi 74% diprediksi benar, rata-rata dari precision dan recall kelas pribadi yaitu 79%.

Model kemudian dilakukan hyperparameter tuning untuk meningkatkan kinerja model. Setelah hyperparameter tuning dengan GridSearchCV 5-fold, akurasi test meningkat menjadi 80,5% (*lihat Tabel VI*), tidak terlihat overfitting pada model. Precision online 86%, Recall online 67%; Precision pribadi 78%, Recall pribadi 91% seperti terlihat pada Tabel VII.

C. Feature Importance

Gambar 1 menunjukkan ranking feature importance yang dianalisis menggunakan Random Forest.

Fitur jarak perjalanan memiliki skor terendah dan kemudian dihapus dalam analisis lanjutan. Setelah menyeleksi fitur terendah, akurasi test meningkat menjadi 85,36% (*lihat tabel VIII*), tidak terlihat overfitting pada model. Precision online



Gambar 1. Hasil feature importance

TABEL VIII. AKURASI MODEL SETELAH SELEKSI FITUR

| Data | Akurasi |
|----------|---------|
| Training | 84% |
| Testing | 85,36% |

TABEL IX. CLASSIFICATION REPORT SETELAH SELEKSI FITUR

| | Precision | Recall | F1-Score |
|---|-----------|--------|----------|
| 0 | 0,93 | 0,72 | 0,81 |
| 1 | 0,81 | 0,96 | 0,88 |

93%, Recall online 72%; Precision pribadi 81%, Recall pribadi 96% seperti terlihat di Tabel IX.

D. Perbandingan Performa Model

Perbandingan performa dari ketiga model yang dilakukan telah dituangkan dalam Tabel X. Model default menampilkan akurasi 100% pada data latih namun hanya 78% pada data uji. Perbedaan besar ini menandakan overfitting, di mana model menghafal data latih dan gagal menggeneralisasi ke data baru. Setelah tuning dengan GridSearchCV, akurasi latih turun menjadi 84,5% sementara akurasi uji meningkat menjadi 80,5%. Artinya, model ter-tuning lebih seimbang dengan selisih akurasi latih–uji yang mengecil, sesuai dengan konsep bahwa selisih besar antara akurasi latih dan uji menunjukkan model *overfit*. Pengaturan hiperparameter (misalnya jumlah pohon, kedalaman maksimum, dll.) terbukti meningkatkan kemampuan generalisasi model.

Peningkatan juga terlihat pada metrik evaluasi lainnya. Precision rata-rata untuk model default sekitar 79% naik menjadi 81%, pada model ter-tuning, dan 86% setelah penghapusan fitur. Recall masing-masing adalah 78%, 80%, dan 85%, sedangkan F1-score (rata-rata harmonik dari precision dan recall) berurutan 78%, 80%, dan 85%. Precision mengukur proporsi prediksi positif yang benar, sedangkan recall mengukur berapa banyak kasus positif aktual yang berhasil dideteksi. F1-score memadukan kedua metrik tersebut sebagai rata-rata harmonik, memberikan indikator keseimbangan model. Kenaikan F1-score yang lebih besar pada model ketiga (85%) menunjukkan perbaikan seimbang antara ter-tuning dengan seleksi fitur tidak hanya lebih akurat secara

TABEL X. PERBANDINGAN PERFORMA MODEL

| Deskripsi Model | Akurasi | | Precision (weighted avg.) | Recall (weighted avg.) | F1-Score (weighted avg.) |
|-----------------------------------|----------|---------|---------------------------|------------------------|--------------------------|
| | Training | Testing | | | |
| Default | 100% | 78% | 79% | 78% | 78% |
| GridSearchCV | 84,5% | 80,5% | 81% | 80% | 80% |
| GridSearchCV Dengan Seleksi Fitur | 84% | 85,3% | 86% | 85% | 85% |

keseluruhan, tetapi juga lebih konsisten dalam mengklasifikasikan moda transportasi online maupun kendaraan pribadi.

Fitur jarak perjalanan memiliki nilai penting terendah dalam model (hampir tidak berkontribusi). Setelah fitur ini dihapus, performa model membaik ditandai dengan akurasi testing melonjak dari 80,5% menjadi 85,4%, dan precision/recall/F1-score masing-masing bertambah sekitar 5 poin persentase (*lihat Tabel X*). Hasil ini menunjukkan bahwa jarak perjalanan semula hanya menambah noise tanpa informasi prediktif signifikan. Menghilangkan variabel yang sangat kurang pengaruhnya ini mengurangi kompleksitas dimensi data tanpa kehilangan sinyal utama, sehingga mengurangi risiko overfitting. Dengan kata lain, model ketiga menjadi lebih sederhana tetapi tetap kuat.

Temuan ini sejalan dengan studi dari [11] bahwa setelah dilakukan tuning dan seleksi fitur pada model maka performa model semakin baik yang ditandai dengan akurasi data test dari 66,5% naik menjadi 70,1%. Hal ini sejalan dengan prinsip bahwa Random Forest cenderung lebih akurat jika difokuskan pada fitur paling relevan.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil penelitian ini, karakteristik pelaku perjalanan didominasi oleh responden berjenis kelamin perempuan, mayoritas responden berusia 20 – 29 tahun, sebagian besar responden berstatus pelajar/mahasiswa, pendapatan umumnya di bawah 2 juta, kebanyakan memiliki sepeda motor, mayoritas responden bertempat tinggal di Malalayang. Karakteristik perjalanan didominasi oleh jarak perjalanan sejauh 1 – 5 km, durasi perjalanan 0 – 15 menit, biaya perjalanan Rp10.000 – 20.000, kunjungan ke Kawasan Megamas Manado mayoritas 1 – 2 kali seminggu, dengan tujuan perjalanan umumnya yaitu rekreasi/hiburan. Berdasarkan karakteristik preferensi, responden cenderung memilih moda dengan pertimbangan waktu, aksesibilitas, dan biaya. Kemudian, berdasarkan pemilihan moda mayoritas responden memilih kendaraan pribadi dibandingkan dengan transportasi online.

Analisis *feature importance* pada model Random Forest

menunjukkan faktor yang mempengaruhi pengambilan keputusan responden dalam memilih moda transportasi yaitu kepemilikan kendaraan pribadi sebesar 25,4%, diikuti oleh alasan utama sebesar 22,3%, usia sebesar 9,4%, tempat tinggal sebesar 9,3%, pekerjaan sebesar 7%, biaya perjalanan sebesar 6,3%, sementara faktor yang lain memiliki nilai tingkat kepentingan di bawah 5%.

Penelitian ini menunjukkan algoritma Random Forest sangat cocok untuk klasifikasi moda transportasi. Hyperparameter tuning menggunakan GridSearchCV menghasilkan model yang lebih baik yaitu 80,5% pada data test. Proses tuning meningkatkan kemampuan generalisasi model dan meminimalkan perbedaan akurasi antara data train dan test. Selanjutnya, dengan menghapus fitur terendah model mencapai akurasi tertinggi 85,36% pada data test, tidak mengalami overfitting, serta recall dan precision yang baik untuk kedua kelas moda. Dengan demikian, Random Forest yang dioptimasi dengan GridSearchCV serta penghapusan fitur tidak penting terbukti dapat diandalkan untuk memprediksi pilihan moda transportasi di kota Manado.

B. Saran

Tarif parkir di Megamas Manado digandakan untuk mobil dengan solo rider, sedangkan mobil berpenumpang ≥ 2 tetap membayar tarif normal. Kemudian, optimalkan fitur ridesharing di aplikasi online agar penumpang berarah serupa dapat berbagi biaya, dan berikan kompensasi (misalnya voucher Rp. 5.000–10.000) jika driver terlambat lebih dari 5 menit dari estimasi. Untuk Random Forest, selain GridSearchCV pertimbangkan RandomizedSearchCV agar eksplorasi ruang parameter lebih cepat dan efisien.

KUTIPAN

- [1] Badan Pusat Statistik Manado, *Kota Manado Dalam Angka*, vol. 21. Manado: BPS Kota Manado, 2024.
- [2] S. Y. R. Rompis, "Karakteristik Pemilihan Moda di Kota Manado Dengan Metode Multinomial Logit," *Jurnal Penelitian Jalan dan Jembatan*, vol. 1, no. 1, 2021, [Online]. Available: <https://doi.org/10.59900/ptrkjj.v1i1.25>
- [3] J. Matulende, L. Lefrandt, and S. Pandey, "Pengaruh Angkutan Online Terhadap Pemilihan Moda Transportasi Publik Di Kota Manado (Studi Kasus: Trayek Sumopmo-Pusat Kota)," *TEKNO*, vol. 20, no. 81, 2022, [Online]. Available: <https://doi.org/10.35793/jts.v20i81.43496>
- [4] R. L. Hermaputi and C. Hua, "Decoding Jakarta Women's Non-Working Travel-Mode Choice: Insights from Interpretable Machine-Learning Models," *Sustainability (Switzerland)*, vol. 16, no. 19, Oct. 2024, doi: 10.3390/su16198454.
- [5] L. Lefrandt and M. Kumaat, "Characteristics of Transportation Mode Selection in Manado Maritime City," *International Journal of Marine Engineering Innovation and Research*, vol. 9, no. 4, pp. 2548–1479, 2024.
- [6] M. Maranatha, M. Audie, L. E. Rumayar, and L. Jefferson, "Model Pemilihan Moda Angkutan Umum dan Transportasi Online di Kota Tomohon (Studi Kasus: Pelajar di Kota Tomohon)," *Jurnal Sipil Statik*, vol. 8, no. 6, pp. 911–924, 2020.
- [7] N. Fahriza *et al.*, "Travel Mode Choice Modeling: Predictive Efficacy between Machine Learning Models and Discrete Choice Model," *The Open Transportation Journal*, vol. 15, 2021, doi: 10.2174/18744478021150102.
- [8] C. R. Sekhar, Minal, and E. Madhu, "Mode Choice Analysis Using Random Forrest Decision Trees," in *Transportation Research Procedia*, Elsevier B.V., 2016, pp. 644–652. doi: 10.1016/j.trpro.2016.11.119.
- [9] H. Zhang, L. Zhang, Y. Liu, and L. Zhang, "Understanding Travel Mode Choice Behavior: Influencing Factors Analysis and Prediction with Machine Learning Method," *Sustainability (Switzerland)*, vol. 15, no. 14, Jul. 2023, doi: 10.3390/su151411414.
- [10] J. J. M. D'Cruz, A. P. Alex, and V. S. Manju, "MODE CHOICE ANALYSIS OF SCHOOL TRIPS USING RANDOM FOREST TECHNIQUE," *Archives of Transport*, vol. 63, no. 3, pp. 39–48, 2022, doi: 10.5604/01.3001.0015.9175.
- [11] N. F. Mohd Ali, A. F. Mohd Sadullah, A. P. P. Abdul Majeed, M. A. Mohd Razman, and R. M. Musa, "The identification of significant features towards travel mode choice and its prediction via optimised random forest classifier: An evaluation for active commuting behavior," *J Transp Health*, vol. 25, Jun. 2022, doi: 10.1016/j.jth.2022.101362.
- [12] H. Zhang and N. Liu, "Behavioral Analysis of Urban Travel Mode Selection Based on Random Forest Algorithm," *International Journal of Multiphysics*, vol. 18, no. 3, 2024.
- [13] H. A. Kalantari, S. Sabouri, S. Brewer, R. Ewing, and G. Tian, "Machine Learning in Mode Choice Prediction as Part of MPOs' Regional Travel Demand Models: Is It Time for Change?," *Sustainability (Switzerland)*, vol. 17, no. 8, Apr. 2025, doi: 10.3390/su17083580.
- [14] L. Breiman, "Random Forests," 2001.
- [15] F. Pedregosa FABIANPEDREGOSA *et al.*, "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011, [Online]. Available: <http://scikit-learn.sourceforge.net>.

Kezia Gracella Najoan lahir di Manado pada 26 November



2001. Penulis menempuh pendidikan dasar di SD Kr. Eben Haezar 1 Manado, kemudian melanjutkan ke SMP Kr. Eben Haezar 1 Manado. Setelah itu ia bersekolah di SMA Negeri 9 Manado. Selanjutnya, penulis melanjutkan studi Sarjana (S1) pada Program Studi Teknik Sipil, Fakultas Teknik, Universitas Sam Ratulangi Manado. Kemudian ia memperdalam bidang ilmunya dengan menempuh pendidikan Magister (S2) di Program Studi Magister Teknik Sipil, Program Pascasarjana, Universitas Sam Ratulangi.