Jurnal JPILKOM, 3 (2) (2025), ISSN: 3025-6887 (Online)
Jurnal JPILKOM (Jurnal Penelitian Ilmu Komputer)
Journal homepage: https://jpilkom.org

Optimasi Hyperparameter Random Forest dalam Memprediksi Daya Beli Mobil Menggunakan GridSearch

Retno Pradistiani
Program Studi Sistem Informasi, STIKOM Tunas Bangsa Pematangsiantar, Indonesia
E-Mail: rpradistiani@gmail.com

Article Info

Article history:

Received Mei 01, 2025
Revised Mei 30, 2025
Accepted Jun 10, 2025

Kata Kunci:

Random Forest
Prediksi Pembelian Mobil
SMOTE
GridSearchCV
Machine Learning

Keywords:

Random Forest
Car Purchase Prediction
SMOTE
GridSearchCV
Machine Learning

ABSTRAK

Prediksi keputusan pembelian kendaraan merupakan topik penting dalam pengembangan strategi pemasaran berbasis data. Penelitian ini bertujuan untuk membangun model machine learning menggunakan algoritma Random Forest guna memprediksi keputusan pembelian mobil berdasarkan data demografis dan ekonomi pengguna. Proses pelatihan model didukung oleh teknik oversampling SMOTE untuk mengatasi ketidakseimbangan kelas, serta optimasi hyperparameter menggunakan GridSearchCV. Hasil penelitian menunjukkan bahwa model Random Forest dengan parameter terbaik (n_estimators = 200, max_depth = 20, min_samples_split = 10, min_samples_leaf = 1, dan max_features = 'sqrt') mampu mencapai akurasi sebesar 93,33%. Berdasarkan classification report, kelas 0 memiliki precision sebesar 0,93 dan recall 0,95, sedangkan kelas 1 memiliki precision 0,94 dan recall 0,92. Model menunjukkan kinerja yang konsisten pada metrik presisi dan recall, yang membuktikan efektivitasnya dalam menangani ketidakseimbangan kelas. Penelitian ini memberikan kontribusi dalam menyediakan solusi prediktif yang andal untuk meningkatkan efisiensi strategi pemasaran dan segmentasi konsumen di industri otomotif dan keuangan.

ABSTRACT

Predicting vehicle purchase decisions is an essential topic in the development of data-driven marketing strategies. This study aims to build a machine learning model using the Random Forest algorithm to predict car purchase decisions based on users' demographic and economic data. The model training process was supported by SMOTE oversampling to address class imbalance and hyperparameter tuning using GridSearchCV. The results show that the best-performing Random Forest model (n_estimators = 200, max_depth = 20, min_samples_split = 10, min_samples_leaf = 1, and max_features = 'sqrt') achieved an accuracy of 93.33%. According to the classification report, class 0 had a precision of 0.93 and recall of 0.95, while class 1 had a precision of 0.94 and recall of 0.92. The model demonstrated consistent performance in precision and recall metrics, confirming its effectiveness in handling class imbalance. This research contributes a reliable predictive solution to support marketing strategy efficiency and consumer segmentation in the automotive and financial industries.

Corresponding Author:
Retno Pradistiani,
Program Studi Sistem Informasi, STIKOM Tunas Bangsa Pematangsiantar,
Jalan Jendral Sudirman Blok A, No. 1/2/3, Siantar Barat, Kota Pematangsiantar, Sumatera Utara, 21127, Indonesia
Email: rpradistiani@gmail.com

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mendorong pemanfaatan metode kecerdasan buatan dalam mendukung proses pengambilan keputusan berbasis data. Salah satu pendekatan yang kini banyak digunakan dalam dunia industri dan keuangan adalah *machine learning*, terutama dalam tugas-tugas prediksi dan klasifikasi. Algoritma *Random Forest*, sebagai salah satu teknik pembelajaran *ensemble* yang populer, telah banyak digunakan dalam berbagai studi karena kemampuannya menghasilkan akurasi tinggi dan ketahanan terhadap *overfitting*.Topik prediksi berbasis *machine learning* menjadi semakin relevan, terutama dalam konteks ekonomi digital yang berkembang pesat. Salah satunya adalah upaya memprediksi daya beli masyarakat, yang sangat penting untuk perusahaan dalam menyusun strategi penjualan dan pemasaran. Beberapa penelitian terdahulu telah membuktikan efektivitas metode ini, termasuk dalam prediksi daya beli mobil dengan pendekatan *Multilayer Perceptron* dan optimasi *hyperparameter* yang mampu meningkatkan akurasi model secara signifikan (Iqbal et al., 2023)

Pemahaman masyarakat dan peneliti terhadap penerapan *machine learning* telah meluas ke berbagai sektor. Dalam bidang kesehatan, algoritma *Random Forest* telah digunakan untuk memprediksi kemungkinan diabetes (Apriliah et al., 2021), penyakit stroke (Dana et al., 2024; Fadli & Saputra, 2023), hingga kanker paru (Sari et al., 2023). Di sektor keuangan dan ekonomi, algoritma ini diterapkan untuk memprediksi harga Bitcoin (Saadah & Salsabila, 2021), kelayakan kredit pinjaman (Prasojo & Haryatmi, 2021), serta harga mobil bekas (Dana et al., 2024; Dewi et al., 2024; Kriswantara & Sadikin, 2022; Winata, 2024). Selain itu, terdapat juga studi yang membandingkan algoritma *Random Forest* dengan metode lain seperti *K-Nearest Neighbor* (KNN) dalam prediksi harga mobil (Sulaiman & Negara, 2023), menunjukkan performa yang kompetitif dari *Random Forest* dalam skenario dunia nyata. Namun, hingga kini masih terdapat sejumlah permasalahan yang belum terselesaikan. Salah satu tantangan utama adalah kebutuhan tuning parameter yang tepat untuk memaksimalkan performa algoritma *Random Forest* dalam konteks data yang spesifik. Selain itu, penggunaan model yang tepat dalam konteks prediksi daya beli masyarakat—khususnya dalam sektor otomotif seperti pembelian mobil—masih jarang dijumpai, terutama dalam konteks geografis dan ekonomi tertentu yang berbeda seperti Indonesia. Beberapa penelitian yang fokus pada prediksi harga mobil menggunakan *Random Forest* atau algoritma lain belum secara eksplisit membahas daya beli sebagai variabel target.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun dan mengevaluasi model prediksi daya beli mobil menggunakan algoritma *Random Forest*. Dengan pendekatan ini, diharapkan dapat memberikan kontribusi dalam bentuk model prediktif yang akurat dan andal dalam membantu perusahaan otomotif, perbankan, dan lembaga keuangan untuk mengidentifikasi potensi pelanggan yang memiliki daya beli terhadap kendaraan. Model ini mengintegrasikan berbagai fitur yang relevan dan melakukan tuning hyperparameter untuk memperoleh performa optimal, sebagaimana telah dikaji dalam beberapa penelitian sebelumnya.

Penelitian ini juga diharapkan dapat memperkaya literatur dan praktik penerapan *machine learning* untuk tujuan klasifikasi dan prediksi di sektor otomotif serta menjadi rujukan bagi penelitian lanjutan dalam pengembangan model yang adaptif terhadap dinamika pasar Indonesia.

2. METODE PENELITIAN

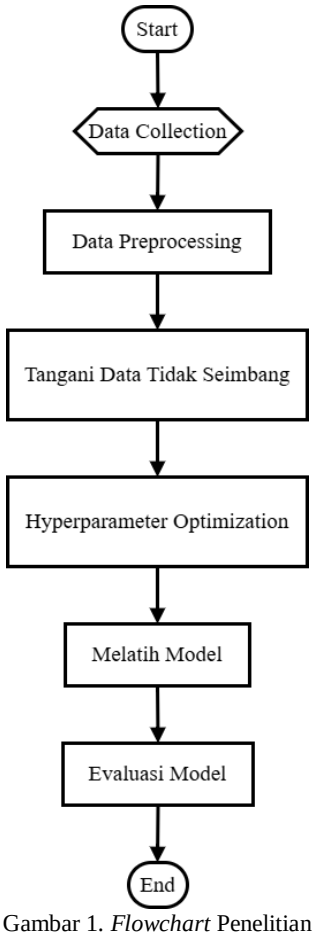
2.1 *Data Penelitian*

Dataset yang digunakan dalam penelitian ini merupakan data keputusan pembelian mobil yang diambil dari *Cars Purchase Decision Dataset* yang tersedia di *platform Kaggle* (Gabriel, 2022). Dataset ini terdiri dari 1.000 entri data dengan lima fitur utama, yaitu: "*User ID*", yang merupakan identitas unik untuk setiap responden; "*Gender*", variabel kategorikal yang menunjukkan jenis kelamin (‘*Male*’ atau ‘*Female*’); "*Age*", usia pengguna dalam bentuk bilangan bulat; "*Annual Salary*", jumlah pendapatan tahunan pengguna; dan "*Purchased*", variabel target biner dengan nilai 0 jika tidak membeli dan 1 jika membeli mobil.

Penelitian ini bertujuan untuk memprediksi keputusan pembelian mobil berdasarkan fitur usia, jenis kelamin, dan penghasilan tahunan menggunakan algoritma *Random Forest*. Untuk meningkatkan akurasi model, dilakukan optimasi hyperparameter menggunakan *GridSearch*. Selain itu, digunakan pula teknik *preprocessing* seperti *oversampling* dengan SMOTE untuk menangani ketidakseimbangan kelas untuk meningkatkan performa prediktif model secara keseluruhan.

2.2 *Flowchart Penelitian*

Penelitian ini mengikuti alur kerja yang terstruktur untuk memastikan setiap langkah ditangani secara sistematis dan guna meningkatkan kinerja model. Gambar 1 menggambarkan proses yang dilakukan secara bertahap dalam penelitian ini.



Gambar 1. *Flowchart* Penelitian

Penelitian ini dimulai dengan tahap awal (Start) yang menandai dimulainya proses pemodelan prediktif terhadap keputusan pembelian mobil. Langkah pertama yang dilakukan adalah pengumpulan data (*Data Collection*), di mana data diambil dari *Cars Purchase Decision Dataset* yang tersedia di platform Kaggle. Dataset ini memuat 1.000 entri dengan lima fitur, yaitu *User ID, Gender, Age, Annual Salary,* dan *Purchased* sebagai variabel target. Setelah data dikumpulkan, dilakukan tahap pra-pemrosesan data (*Data Preprocessing*). Pada tahap ini, data dibersihkan dan disiapkan untuk pelatihan model, termasuk pemilihan fitur yang relevan seperti *Gender*, *Age,* dan *Annual Salary*. Selanjutnya, dilakukan penanganan ketidakseimbangan kelas (Tangani Data Tidak Seimbang) karena distribusi antara data pembeli dan non-pembeli tidak seimbang. Untuk itu, digunakan teknik *oversampling* seperti SMOTE guna menyeimbangkan distribusi kelas pada data latih.

Kemudian, model mengalami tahap optimasi hyperparameter (*Hyperparameter Optimization*) dengan menggunakan metode *GridSearch* untuk menemukan kombinasi parameter terbaik bagi algoritma *Random Forest*. Setelah parameter terbaik diperoleh, model dilatih pada data yang telah diproses dalam tahap pelatihan model (Melatih Model). Model ini belajar mengenali pola-pola yang berkaitan dengan keputusan pembelian mobil berdasarkan fitur-fitur yang disediakan. Langkah berikutnya adalah evaluasi model (Evaluasi Model), di mana performa model diuji menggunakan metrik evaluasi yang sesuai untuk mengukur akurasi dan efektivitas dalam memprediksi keputusan pembelian. Akhir dari proses ini ditandai dengan tahap selesai (*End*), di mana model telah selesai dibangun dan dievaluasi, siap untuk digunakan dalam pengambilan keputusan atau implementasi lebih lanjut.

2.3 *Random Forest*

Beberapa studi menunjukkan bahwa penggunaan *Random Forest* memberikan hasil yang kuat dalam berbagai tugas klasifikasi, terutama ketika diterapkan pada data berdimensi tinggi dan kompleks (Salman et

al., 2024). Dalam bidang penginderaan jauh, algoritma ini menghasilkan akurasi yang kompetitif jika dibandingkan dengan metode lain seperti *Support Vector Machine*, terutama saat menangani jumlah fitur yang besar dan data yang tidak linier (Sheykhmousa et al., 2020). Selain itu, penerapannya juga ditemukan efektif dalam konteks interpolasi spasial, di mana *Random Forest* mampu menangkap variasi spasial secara akurat tanpa bergantung pada asumsi statistik yang ketat (Sekulić et al., 2020). Ketiga studi tersebut memperkuat posisi algoritma ini sebagai metode yang fleksibel dan memiliki performa prediktif yang baik di berbagai bidang.

2.4 *Oversampling dengan SMOTE*

SMOTE digunakan untuk menyeimbangkan jumlah data pada setiap kelas variabel target. Proses dilakukan setelah *encoding* dan sebelum pembagian data menjadi data latih dan data uji, agar model tidak belajar dari data uji. Pemilihan SMOTE didasarkan pada hasil penelitian sebelumnya yang menunjukkan bahwa metode ini efektif meningkatkan kinerja klasifikasi pada data tidak seimbang, baik pada model *Random Forest*, CNN, maupun kombinasi dengan metode lain (Joloudari et al., 2023)(Nguyen et al., 2023)(Nurdian et al., 2022)(Salman et al., 2024)(Wongvorachan et al., 2023).

2.5 *Hyperparameter Tuning*

Tuning hyperparameter dilakukan menggunakan metode *GridSearch* untuk memperoleh kombinasi parameter terbaik pada model *Random Forest*. Proses ini dilakukan dengan mengatur beberapa nilai kandidat untuk parameter seperti *n_estimators*, *max_depth*, *min_samples_split*, dan *min_samples_leaf*. Pemilihan *GridSearch* didasarkan pada keunggulannya dalam menjelajahi seluruh ruang kombinasi parameter secara sistematis, sehingga mampu meningkatkan akurasi model secara signifikan (Nurdian et al., 2022). Selain itu, *GridSearch* juga efektif meningkatkan performa klasifikasi sentimen menggunakan *Random Forest*, dengan hasil akurasi yang lebih stabil dibandingkan metode tuning lainnya (Siji George & Sumathi, 2020). Dalam penelitian ini, *GridSearch* dikombinasikan dengan validasi silang untuk memastikan model yang diperoleh memiliki performa yang konsisten dan tidak *overfitting*.

2.6 *Evaluasi Model*

Evaluasi model dilakukan untuk mengukur kinerja *Random Forest* dalam memprediksi daya beli mobil. Beberapa metrik yang digunakan meliputi akurasi, *precision*, *recall*, dan *F1-Score*. Seluruh metrik dievaluasi pada data uji untuk memastikan kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Akurasi digunakan sebagai metrik utama karena mencerminkan proporsi prediksi yang benar terhadap keseluruhan data. Namun, untuk memastikan evaluasi yang seimbang terhadap masing-masing kelas, metrik *precision*, *recall*, dan *F1-Score* juga dianalisis secara per kelas. Selain itu, digunakan *confusion matrix* untuk melihat distribusi prediksi model pada masing-masing kelas. Evaluasi dilakukan setelah proses tuning hyperparameter agar hasil mencerminkan performa terbaik dari model yang telah dioptimasi.

3. HASIL DAN PEMBAHASAN

3.1 *Data Collection*

Data yang digunakan dalam penelitian ini diperoleh dari hasil survei daya beli konsumen terhadap kendaraan. Dataset terdiri dari berbagai fitur seperti ID User, jenis kelamin, usia, dan pendapatan Tahunan yang dapat memengaruhi keputusan pembelian mobil. Data telah dikumpulkan dalam bentuk file terstruktur dan dimuat menggunakan perangkat lunak pengolah data untuk analisis lebih lanjut. Gambar 2 menunjukkan 5 baris pertama dari dataset yang digunakan.

	User ID	Gender	Age	AnnualSalary	Purchased
0	385	Male	35	20000	0
1	681	Male	40	43500	0
2	353	Male	49	74000	0
3	895	Male	40	107500	1
4	661	Male	25	79000	0

Gambar 2. Beberapa Baris Dataset

Dari dataset yang telah dikumpulkan tidak terdapat baris ataupun kolom yang berisi *missing values*, duplikasi data dan *outliers*. Sehingga dapat dilanjutkan ke proses berikutnya.

3.2 *Data Preprocessing*

Proses *preprocessing* dimulai dengan menghapus kolom (*feature selection*) yang tidak memiliki kontribusi terhadap prediksi, seperti kolom *User ID*. Selanjutnya, fitur kategorikal dikonversi menjadi numerik menggunakan teknik *Label Encoding*, pada dataset tersebut hanya kolom *Gender* yang merupakan fitur kategorikal, dan hanya memiliki 2 variasi data yaitu *Male* dan *Female*, sehingga ketika diencode akan menghasilkan 0 dan 1, yang mana 0 untuk *Male* dan 1 untuk *Female*. Gambar 3 menunjukkan dataset yang sudah bersih dari *variable* yang tidak dibutuhkan dan yang telah dilakukan *encoding* dengan teknik *Label Encoding*.

	Gender	Age	AnnualSalary	Purchased
0	1	35	20000	0
1	1	40	43500	0
2	1	49	74000	0
3	1	40	107500	1
4	1	25	79000	0

Gambar 3. Data *Feature Selection* dan Hasil *Encoding*

Normalisasi juga diterapkan pada fitur numerik untuk menyamakan skala antar variabel. Proses ini penting untuk memastikan setiap fitur memiliki kontribusi yang seimbang terhadap model. Gambar 4 menunjukkan data bersih setelah dilakukan normalisasi data.

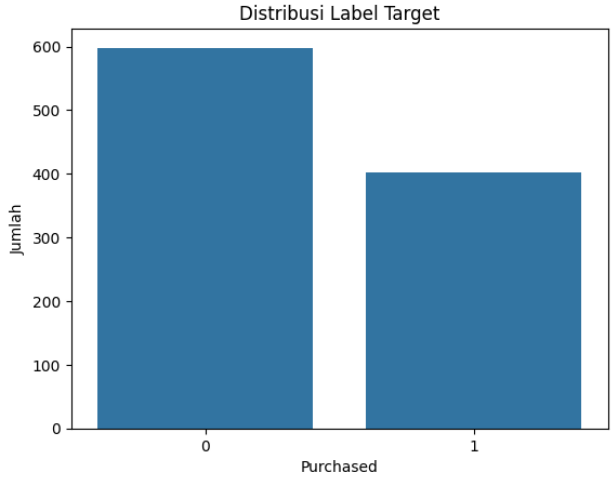
	Gender	Age	AnnualSalary	Purchased
0	1	0.377778	0.036364	0
1	1	0.488889	0.207273	0
2	1	0.688889	0.429091	0
3	1	0.488889	0.672727	1
4	1	0.155556	0.465455	0

Gambar 4. Data Hasil Normalisasi Data

3.3 *Tangani Data Tidak Seimbang*

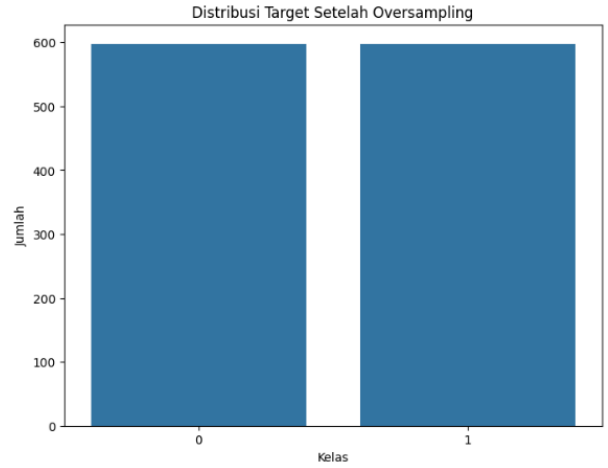
Setelah data dinormalisasi maka data dapat pisahkan fitur dan targetnya. Pada dataset ini yang merupakan fitur adalah *Age*, *Gender* dan *AnnualSalary*. Sedangkan Target adalah *Purchased*, dengan jumlah 2 variasi kelas, yaitu 0 dan 1, yang mana 0 artinya tidak membeli dan 1 artinya membeli. Distribusi kelas pada variabel target awalnya tidak seimbang, yaitu distribusi pada kelas 0 ada sebanyak 598 data dan pada kelas 1

ada sebanyak 402 data, yang dapat menyebabkan bias prediksi pada model. Gambar 5 menunjukkan visualisasi data sebelum dilakukan penangangan distribusi kelas yang tidak seimbang.



Gambar 5. Distribusi Kelas Sebelum *Oversampling*

Oleh karena itu, digunakan metode SMOTE untuk melakukan *oversampling* terhadap kelas minoritas. SMOTE menghasilkan data sintetis berdasarkan hubungan antar sampel di kelas minoritas, sehingga model dapat belajar secara adil terhadap semua kelas. Gambar 6 menunjukkan distribusi kelas setelah dilakukan *oversampling*.



Gambar 6. Distribusi Kelas Setelah *Oversampling*

3.4 *Hyperparameter Tuning*

Setelah data sudah dilakukan penyeimbangan terhadap distribusi data, maka selanjutnya data dibagi menjadi data train dan data test, dengan proporsi 80% untuk data *training* dan 20% untuk data *testing*. Untuk memperoleh kinerja optimal dari model *Random Forest*, dilakukan tuning *hyperparameter* menggunakan metode *GridSearch* dengan validasi silang sebanyak 5 *fold* terhadap 384 kombinasi parameter, sehingga total dilakukan 1920 proses pelatihan. Parameter yang diuji meliputi jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), jumlah minimum sampel untuk split (*min_samples_split*), jumlah minimum sampel di daun (*min_samples_leaf*), dan jumlah fitur pada setiap split (*max_features*). Hasil terbaik diperoleh pada kombinasi parameter: *n_estimators* sebanyak 200, *max_depth* 20, *min_samples_split* 10, *min_samples_leaf* 1, dan *max_features* menggunakan nilai *sqrt*.

3.5 *Melatih Model*

Model *Random Forest* dilatih menggunakan data latih yang telah melalui proses *oversampling* dengan SMOTE. Pelatihan dilakukan menggunakan kombinasi *hyperparameter* terbaik yang diperoleh dari *GridSearch*. Selama proses pelatihan, kinerja model dimonitor menggunakan metrik evaluasi terhadap data validasi untuk memastikan kestabilan dan menghindari *overfitting*.

3.6 *Evaluasi Model*

Evaluasi dilakukan pada data uji menggunakan metrik akurasi, *precision, recall*, dan *F1-score* untuk mengukur performa model dalam mengklasifikasikan daya beli konsumen. Hasil evaluasi menunjukkan bahwa model *Random Forest* mampu memberikan kinerja yang sangat baik dengan akurasi mencapai **93,33%.** Nilai *precision, recall*, dan *F1-score* pada masing-masing kelas tergolong tinggi dan seimbang, mengindikasikan bahwa model tidak berat sebelah terhadap salah satu kelas. Gambar 7 menunjukkan *Classification Report* berdasarkan model yang dibangun menggunakan parameter terbaik dari *hyperparameter tuning* yang sudah dilakukan.

Classification Report:

	precision	recall	f1-score	support
0	0.93	0.95	0.94	129
1	0.94	0.92	0.93	111
accuracy			0.93	240
macro avg	0.93	0.93	0.93	240
weighted avg	0.93	0.93	0.93	240

Gambar 7. *Classification Report*

Berdasarkan *Classification Report*, kelas 0 memiliki *precision* sebesar 0,93 dan *recall* 0,95, sedangkan kelas 1 memiliki *precision* 0,94 dan *recall* 0,92. Hal ini menunjukkan bahwa model mampu mengidentifikasi kedua kelas dengan akurasi yang konsisten. Gambar 8 menunjukkan Hasil *Confusion Matrix* dari model yang sudah dilakukan.

```
Confusion Matrix:
[[122   7]
 [  9 102]]
```

Gambar 8. *Confusion Matrix*

Confusion matrix menunjukkan bahwa dari 129 sampel kelas 0, sebanyak 122 diklasifikasikan dengan benar dan hanya 7 salah klasifikasi. Sementara dari 111 sampel kelas 1, sebanyak 102 diklasifikasikan dengan benar dan 9 salah klasifikasi.

4. KESIMPULAN

Penelitian ini berhasil membangun model prediksi keputusan pembelian mobil menggunakan algoritma *Random Forest* dengan dukungan teknik *preprocessing* yang sistematis, *oversampling* dengan SMOTE, dan optimasi hyperparameter menggunakan *GridSearch*. Berdasarkan hasil evaluasi, model yang dikembangkan menunjukkan performa klasifikasi yang baik, dengan nilai akurasi, *precision, recall*, dan *F1-score* yang tinggi dan seimbang pada kedua kelas target.

Adapun kesimpulan yang dapat diambil dari penelitian ini adalah:

1. *Algoritma Random Forest* terbukti efektif dalam memprediksi keputusan pembelian mobil berdasarkan fitur usia, jenis kelamin, dan pendapatan tahunan pengguna.
2. Teknik SMOTE berhasil menangani ketidakseimbangan kelas pada data pelatihan, sehingga meningkatkan kemampuan generalisasi model terhadap data baru.
3. *GridSearch* sebagai metode tuning hyperparameter memberikan peningkatan performa yang signifikan, menghasilkan model dengan parameter optimal yang stabil dan tidak *overfitting*.

4. Evaluasi model menunjukkan bahwa pendekatan ini dapat digunakan sebagai alat bantu pengambilan keputusan bagi industri otomotif, perbankan, atau lembaga keuangan dalam mengidentifikasi calon konsumen potensial berdasarkan karakteristik demografis dan ekonomi.
5. Penelitian ini juga membuka peluang untuk pengembangan model serupa dengan skala data yang lebih besar, fitur tambahan, atau penerapan pada konteks geografis lainnya, khususnya di Indonesia.

Dengan demikian, penerapan machine learning khususnya Random Forest dalam konteks prediksi daya beli kendaraan menunjukkan potensi besar untuk diimplementasikan pada sektor industri berbasis data.

REFERENCES

Apriliah, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. *Sistemasi*, *10*(1), 163. https://doi.org/10.32520/stmsi.v10i1.1129

Dana, A. R., Kristananda, R. V., Bagas, M., Wibowo, S., & Prasetya, D. A. (2024). *Perbandingan Algoritma Decision Tree dan Random Forest dengan Hyperparameter Tuning dalam Mendeteksi Penyakit Stroke*. *4*, 66–75.

Dewi, B. E. S., Haikal, S., Sulistyowati, H. ., Fitriani, R., & K, D. P. (2024). Penerapan Machine Learning Menggunakan Algoritma Random Forest Untuk Prediksi. *Jurnal TRIDI*, *2*(1), 20–31.

Fadli, M., & Saputra, R. A. (2023). Klasifikasi Dan Evaluasi Performa Model Random Forest Untuk Prediksi Stroke. *JT: Jurnal Teknik*, *12*(02), 72–80. http://jurnal.umt.ac.id/index.php/jt/index

Gabriel, S. (2022). *Cars - Purchase Decision Dataset*. Kaggle. https://www.kaggle.com/datasets/gabrielsantello/cars-purchase-decision-dataset

Iqbal, M., Hendri, M. N., Ramadhan Saelan, M. R., Sony Maulana, M., Yudhistira, & Mustopa, A. (2023). Optimasi Hyperparameter Multilayer Perceptron Untuk Prediksi Daya Beli Mobil. *Jurnal Manajemen Informatika Dan Sistem Informasi*, *6*(1), 73–81. https://doi.org/10.36595/misi.v6i1.739

Joloudari, J. H., Marefat, A., Nematollahi, M. A., Oyelere, S. S., & Hussain, S. (2023). Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks. *Applied Sciences (Switzerland)*, *13*(6). https://doi.org/10.3390/app13064006

Kriswantara, B., & Sadikin, R. (2022). Used Car Price Prediction with Random Forest Regressor Model. *Journal of Information Systems, Informatics and Computing Issue Period*, *6*(1), 40–49. https://doi.org/10.52362/jisicom.v6i1.752

Nguyen, T., Mengersen, K., Sous, D., & Liquet, B. (2023). SMOTE-CD: SMOTE for compositional data. *PLoS ONE*, *18*(6 June), 1–19. https://doi.org/10.1371/journal.pone.0287705

Nurdian, R. A., Mujib Ridwan, & Ahmad Yusuf. (2022). Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(1), 24–32. https://doi.org/10.28932/jutisi.v8i1.4004

Prasojo, B., & Haryatmi, E. (2021). Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *7*(2), 79–89. https://doi.org/10.25077/teknosi.v7i2.2021.79-89

Saadah, S., & Salsabila, H. (2021). Prediksi Harga Bitcoin Menggunakan Metode Random Forest. *Jurnal Komputer Terapan*, *7*(1), 24–32. https://doi.org/10.35143/jkt.v7i1.4618

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79. https://doi.org/10.58496/bjml/2024/007

Sari, L., Romadloni, A., & Listyaningrum, R. (2023). Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest. *Infotekmesin*, *14*(1), 155–162. https://doi.org/10.35970/infotekmesin.v14i1.1751

Sekulić, A., Kilibarda, M., Heuvelink, G. B. M., Nikolić, M., & Bajat, B. (2020). Random forest spatial interpolation. *Remote Sensing*, *12*(10), 1–29. https://doi.org/10.3390/rs12101687

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 6308–6325. https://doi.org/10.1109/JSTARS.2020.3026724

Siji George, C. G., & Sumathi, B. (2020). Grid search tuning of hyperparameters in random forest classifier for customer feedback sentiment prediction. *International Journal of Advanced Computer Science and Applications*, *11*(9), 173–178. https://doi.org/10.14569/IJACSA.2020.0110920

Sulaiman, & Negara, E. S. (2023). Komparasi Algoritma K-Nearest Neighbors dan Random Forest Pada Prediksi Harga Mobil Bekas. *Jurnal JUPITER*, *15*(1), 337–346. www.cardekho.com.

Winata, A. J. (2024). Prediksi Harga Mobil Bekas Menggunakan Algoritma Gradient Boosting Machine Dan Random Forest. *Jurnal Inovasi Pendidikan*, *6*(1), 52–61. https://journalpedia.com/1/index.php/jip/article/view/1285

Wongvorachan, T., He, S., & Bulut, O. (2023). A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. *Information (Switzerland)*, *14*(1). https://doi.org/10.3390/info14010054