

EVALUASI KINERJA MODEL KLASIFIKASI PEMBELIAN SEPEDA MENGGUNAKAN ALGORITMA RANDOM FOREST

Rizki Rizkyatul Basir¹, Yunita Endra Megiati², Noor Komari Pratiwi³

Universitas Indraprasta PGRI

Jalan Nangka Raya, C Jl. TB Simatupang No.58, Tj. Bar., Kec. Jagakarsa, Jakarta Selatan
[1rizkyatulbasir@gmail.com](mailto:rizkyatulbasir@gmail.com), [2yunitaendra@gmail.com](mailto:yunitaendra@gmail.com), [3noorkomaripratiwi01@gmail.com](mailto:noorkomaripratiwi01@gmail.com)

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi kinerja model klasifikasi pembelian sepeda menggunakan algoritma Random Forest. Perusahaan sering kali menghadapi kendala dalam memahami faktor-faktor yang memengaruhi keputusan pembelian pelanggan. Masalah ini penting untuk diselesaikan agar strategi pemasaran dapat lebih terarah dan efektif. Data yang digunakan mencakup informasi demografi pelanggan seperti usia, penghasilan, status pekerjaan, dan kategori pendidikan. Proses penelitian meliputi pengumpulan data, pembersihan data, pembuatan model, dan evaluasi kinerja model. Hasil penelitian menunjukkan bahwa model Random Forest menghasilkan akurasi sebesar 78%, dengan precision dan recall yang seimbang. Faktor-faktor seperti pendapatan, usia, dan jumlah anak ditemukan memiliki pengaruh signifikan terhadap keputusan pembelian sepeda. Penelitian ini menunjukkan bahwa Random Forest dapat menjadi alat yang efektif untuk mendukung pengambilan keputusan strategis dalam pemasaran sepeda.

Kata Kunci: Random Forest, klasifikasi, pembelian sepeda, evaluasi model

ABSTRACT

This study aims to evaluate the performance of a bicycle purchase classification model using the Random Forest algorithm. Companies often face challenges in understanding the factors that influence customers' purchasing decisions. Addressing this issue is essential to ensure that marketing strategies are more targeted and effective. The data used in this research include demographic information about customers, such as age, income, employment status, and education level. The research process involves data collection, data cleaning, model development, and model performance evaluation. The results show that the Random Forest model achieves an accuracy of 78%, with balanced precision and recall. Factors such as income, age, and the number of children were found to significantly influence bicycle purchasing decisions. This study demonstrates that Random Forest can serve as an effective tool to support strategic decision-making in bicycle marketing.

Keywords: Random Forest, classification, bicycle purchase, model evaluation.

PENDAHULUAN

Perusahaan retail sepeda menghadapi tantangan dalam memahami perilaku pembelian pelanggan yang dipengaruhi oleh berbagai faktor, seperti usia, penghasilan, dan jumlah anggota keluarga. Random Forest adalah algoritma ensemble yang mampu menangani data dengan kompleksitas tinggi dan menghasilkan prediksi yang stabil (Zhou, 2023). Metode ini menggabungkan beberapa pohon keputusan untuk mengurangi risiko overfitting dan meningkatkan akurasi model. Dalam konteks penelitian ini, algoritma Random Forest digunakan untuk memprediksi keputusan pembelian sepeda berdasarkan data pelanggan, yang mencakup informasi demografi seperti usia, pendapatan, dan status pekerjaan.

Penelitian ini bertujuan untuk mengevaluasi kinerja model klasifikasi Random Forest dalam memprediksi keputusan pembelian

sepeda dan mengidentifikasi faktor-faktor utama yang memengaruhi keputusan tersebut. Urgensi penelitian ini terletak pada kebutuhan perusahaan untuk memahami perilaku pelanggan secara mendalam guna mendukung pengambilan keputusan strategis berbasis data, sehingga mampu meningkatkan efektivitas dan efisiensi strategi pemasaran.

METODE PENELITIAN

Penelitian ini menggunakan Bahasa pemrograman Python dengan Jupyter Notebook sebagai instrumennya. Mulai dari pembersihan data hingga pembuatan model, dilakukan menggunakan pustaka Python seperti pandas, scikit-learn, dan matplotlib.

Dataset Head:

	ID	Marital Status	Gender	Income	Children	Education
0	12496	Married	Female	40000	1	Bachelors
1	24107	Married	Male	30000	3	Partial College
2	14177	Married	Male	80000	5	Partial College
3	24381	Single	Male	70000	0	Bachelors
4	25597	Single	Male	30000	0	Bachelors

	Occupation	Home Owner	Cars	Commute	Distance	Region	Age
0	Skilled Manual	Yes	0	0-1 Miles	Europe	42	
1	Clerical	Yes	1	0-1 Miles	Europe	43	
2	Professional	No	2	2-5 Miles	Europe	60	
3	Professional	Yes	1	5-10 Miles	Pacific	41	
4	Clerical	No	0	0-1 Miles	Europe	36	

	Purchased Bike
0	No
1	No
2	No
3	Yes
4	Yes

Gambar 1. Gambar Tabel Dataset

Gambar di atas adalah cuplikan dataset yang digunakan. Dataset lengkap terdiri dari 1.000 entri dan 13 atribut, dengan kolom "Purchased Bike" sebagai target/label untuk klasifikasi. Penelitian ini dilakukan dengan tahapan sebagai berikut:

1. **Pengumpulan Data:** Data transaksi pelanggan dikumpulkan dari dataset yang didapatkan dari platform dataset dan berisi 1.000 entri, mencakup atribut seperti usia, penghasilan, jenis kelamin, status pernikahan, dan kategori pendidikan. Pada tahap ini, dataset memiliki N entri dan m atribut.

Rumus yang digunakan untuk menghitung jumlah data:

$$N = \text{Jumlah dataset} \quad (1)$$

2. **Pembersihan Data:** Menghilangkan kolom yang tidak relevan (seperti ID) dan menangani atribut kategorikal menggunakan teknik encoding untuk memastikan kompatibilitas dengan algoritma Random Forest.

Rumus untuk menghapus kolom yang tidak relevan:

$$X' = N \times (m - k) \quad (2)$$

Encoding kolom kategorikal:

$$X'' = X' + i = 1 \sum k(ci - 1) \quad (3)$$

3. **Pembuatan Model:** Dataset dibagi menjadi data latih (80%) dan data uji (20%). Algoritma Random Forest diterapkan untuk membangun model klasifikasi. Proporsi pembagian biasanya ditentukan dengan parameter p (data latih) dan $1 - p$ (data uji):

$$\text{Data Latih} = N \times p \quad (4)$$

$$\text{Data Uji} = N \times (1 - p) \quad (5)$$

4. **Evaluasi Model:** Kinerja model diukur menggunakan metrik akurasi, precision, recall, dan f1-score untuk memastikan model memiliki performa yang andal. Evaluasi model dilakukan menggunakan metrik:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Akurasi} = \frac{\text{Total Data}}{\text{Jumlah Prediksi Benar}} \quad (7)$$

$$\text{Recall} = TP + FNTP \quad (8)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} + \text{Recall}}{\text{Precision} \times \text{Recall}} \quad (9)$$

HASIL DAN PEMBAHASAN

penelitian menunjukkan bahwa model Random Forest memiliki kinerja yang baik dengan akurasi sebesar 78%. Berikut adalah laporan evaluasi model:

Akurasi Model Random Forest: 0.78 (77.50%)

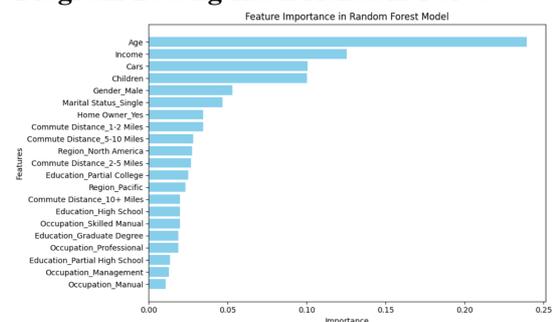
Laporan Klasifikasi:

	precision	recall	f1-score	support
No	0.78	0.80	0.79	106
Yes	0.77	0.74	0.76	94
accuracy			0.78	200
macro avg	0.77	0.77	0.77	200
weighted avg	0.77	0.78	0.77	200

Gambar 2. Gambar Tabel Evaluasi Model

Faktor-faktor yang memiliki pengaruh signifikan terhadap pembelian sepeda adalah pendapatan, usia, dan jumlah anak. Model menunjukkan bahwa pelanggan dengan pendapatan tinggi lebih cenderung membeli sepeda. Usia juga memainkan peran penting, dengan pelanggan yang lebih muda lebih sering membeli sepeda dibandingkan pelanggan yang lebih tua.

Diagram Batang Hasil Random Forest:



Gambar 3. Model Random Forest

Gambar di atas menunjukkan bahwa pendapatan memiliki pengaruh paling signifikan dalam menentukan keputusan pembelian sepeda, diikuti oleh usia dan jumlah anak.

Diskusi: Hasil ini menunjukkan bahwa Random Forest mampu menangkap pola penting dalam dataset. Namun, untuk meningkatkan kinerja model, optimasi lebih lanjut melalui teknik hyperparameter tuning dapat dilakukan, seperti mengatur jumlah pohon (*n_estimators*) atau kedalaman maksimum pohon (*max_depth*).

SIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa model Random Forest dapat digunakan secara efektif untuk memprediksi pembelian sepeda berdasarkan data demografi pelanggan. Model ini menunjukkan kinerja yang cukup baik dengan akurasi sebesar 78%, yang mengindikasikan bahwa algoritma ini mampu menangkap pola utama dalam dataset, termasuk pentingnya fitur seperti pendapatan, usia, dan jumlah anak. Analisis lebih lanjut dari laporan klasifikasi menunjukkan bahwa model memiliki *precision* dan *recall* yang seimbang untuk kedua kelas (*Yes* dan *No*), dengan *f1-score* rata-rata sebesar 77%.

Namun, terdapat beberapa kelemahan yang perlu diperhatikan. Meskipun performa model memadai, terdapat peluang untuk meningkatkan hasil dengan menangani bias kecil yang tampak pada *recall* untuk kelas *Yes*, yang sedikit lebih rendah dibandingkan dengan kelas *No*. Hal ini dapat disebabkan oleh distribusi data yang tidak sepenuhnya seimbang.

Penelitian selanjutnya disarankan untuk mengeksplorasi kombinasi metode *ensemble* lainnya seperti *Gradient Boosting* atau *XGBoost*. Pendekatan ini berpotensi meningkatkan akurasi dan ketahanan model terhadap variasi data. Selain itu, penerapan teknik *hyperparameter tuning* pada model Random Forest, seperti mengoptimalkan jumlah pohon (*n_estimators*) dan kedalaman maksimum (*max_depth*), juga dapat dilakukan untuk meningkatkan kinerja model secara keseluruhan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Indraprasta PGRI dan pihak-pihak

yang telah memfasilitasi dan mendukung penelitian ini dengan data dan fasilitas yang diperlukan.

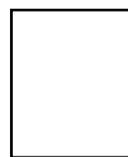
DAFTAR PUSTAKA

- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227. <https://doi.org/10.1007/s11749-016-0481-7> Air. *Jurnal Kefarmasian Indonesia*, 3, 77–88.
- Geurts, P., Ernst, D., & Wehenkel, L. (2023). Extremely randomized trees. *Machine Learning*, 63(1), 3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- Khoshgoftaar, T. M., & Van Hulse, J. (2023). A comprehensive survey of decision tree ensembles: Random Forests, gradient boosting, and beyond. *Journal of Big Data*, 10, 45. <https://doi.org/10.1186/s40537-023-00688-x>
- Pedregosa, F., et al. (2023). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 24(1), 2825–2830. Retrieved from <https://scikit-learn.org/stable>
- Han, J., & Pei, J. (2022). *Data Mining: Concepts and Techniques*. Elsevier.
- Zhou, Z.-H. (2023). *Ensemble Methods: Foundations and Algorithms*. CRC Press.

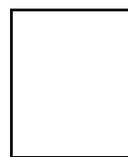
Biografi Penulis



Biografi Penulis 1 Rizki Rizkyatul Basir, Universitas Indraprasta PGRI, Magister Ilmu Komputer, Software Engineering atau Machine Learning



Biografi Penulis 2 Yunita Endra Megiati, Universitas Indraprasta PGRI, Magister Pendidikan, kewarganegaraan dan perubahan sosial



Biografi Penulis 3 Noor Komari Pratiwi, Universitas Indraprasta PGRI, Magister Pendidikan, komunikasi dan literasi gizi