

**PENERAPAN ALGORITMA RANDOM FOREST UNTUK
KLASIFIKASI PELANGGAN BERDASARKAN DATA
DEMOGRAFI DAN PERILAKU BELANJA**

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun Oleh
MIRANDA ALIF RODLIANA
09020222036

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2026

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : MIRANDA ALIF RODLIANA
NIM : 09020222036
Program Studi : Matematika
Angkatan : 2022

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "PENERAPAN ALGORITMA RANDOM FOREST UNTUK KLASIFIKASI PELANGGAN BERDASARKAN DATA DEMOGRAFI DAN PERILAKU BELANJA". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 15 Juni 2026

Yang menyatakan,



MIRANDA ALIF RODLIANA

NIM. 09020222036

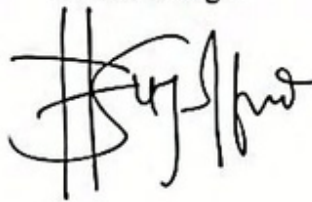
LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : MIRANDA ALIF RODLIANA
NIM : 09020222036
Judul skripsi : PENERAPAN ALGORITMA RANDOM
FOREST UNTUK KLASIFIKASI PELANGGAN
BERDASARKAN DATA DEMOGRAFI DAN
PERILAKU BELANJA

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



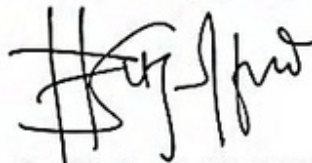
Dr. Yuniar Farida, M.T.
NIP. 197905272014032002

Pembimbing II



Ahmad Lubab, M.Si.
NIP. 198111182009121003

Mengetahui,
Ketua Program Studi Matematika
UIN Sunan Ampel Surabaya



Dr. Yuniar Farida, M.T.
NIP. 197905272014032002

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama : MIRANDA ALIF RODLIANA
NIM : 09020222036
Judul Skripsi : PENERAPAN ALGORITMA RANDOM
FOREST UNTUK KLASIFIKASI PELANGGAN
BERDASARKAN DATA DEMOGRAFI DAN
PERILAKU BELANJA

Telah dipertahankan di depan Tim Penguji
pada tanggal 15 Juni 2026

Mengesahkan,
Tim Penguji

Penguji I



Dr. Dian Candra Rini Novitasari, M. Kom
NIP. 198511242014032001

Penguji II



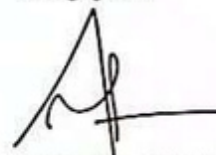
Dr. Abdulloh Hamid, M. Pd
NIP. 198508282014031003

Penguji III



Dr. Yuniar Farida, M.T
NIP. 197905272014032002

Penguji IV



Ahmad Lubab, M.Si
NIP. 198111182009121003

Mengetahui,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya



Dr. Husein Hamdani, M.Pd
NIP. 196507312000031002



UIN SUNAN AMPEL
SURABAYA

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax. 031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Miranda Alif Rodliana
NIM : 09020222036
Fakultas/Jurusan : Sains dan Teknologi / Matematika
E-mail address : mirandaar677@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

☒ Skripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)
yang berjudul :

Penerapan Algoritma Random forest untuk Klarifikasi Pelanggan
Berdasarkan Data Demografi dan Perilaku Belanja

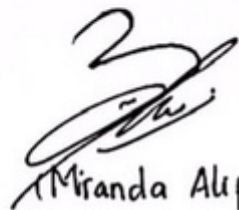
beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara **fulltext** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 15 Juni 2026

Penulis


(Miranda Alif Rodliana)

ABSTRAK

PENERAPAN ALGORITMA RANDOM FOREST UNTUK KLASIFIKASI PELANGGAN BERDASARKAN DATA DEMOGRAFI DAN PERILAKU BELANJA

Perkembangan dunia bisnis yang semakin kompetitif menuntut perusahaan untuk memahami karakteristik pelanggan secara lebih mendalam agar strategi pemasaran dapat disusun secara tepat sasaran. Pemanfaatan data demografi dan perilaku belanja dapat digunakan untuk mengklasifikasikan pelanggan berdasarkan respons terhadap promosi. Penelitian ini membahas penerapan algoritma *Random Forest* dalam klasifikasi pelanggan serta mengevaluasi performa model dalam membedakan pelanggan yang merespons dan tidak merespons promosi. Data yang digunakan adalah *Customer Personality Analysis* yang diperoleh dari Kaggle, berisi variabel demografi seperti usia, pendapatan, tingkat pendidikan, status pernikahan, dan jumlah anak, serta variabel perilaku belanja seperti jumlah pembelian produk, frekuensi transaksi, dan respons terhadap promosi. Penanganan ketidakseimbangan kelas dilakukan menggunakan SMOTENC karena data memiliki kombinasi fitur numerik dan kategorik. Hasil penelitian menunjukkan bahwa model *Random Forest* dengan SMOTENC memperoleh nilai akurasi sebesar 0,863267, presisi sebesar 0,539683, sensitivitas sebesar 0,612613, dan skor F-1 sebesar 0,573840. Penelitian ini diharapkan dapat memberikan gambaran mengenai pemanfaatan algoritma *Random Forest* dalam klasifikasi pelanggan serta mendukung penyusunan strategi pemasaran yang lebih tepat sasaran.

Kata kunci: Data Demografi, Klasifikasi Pelanggan, Perilaku Belanja, *Random Forest*, SMOTENC.

ABSTRACT

IMPLEMENTATION OF RANDOM FOREST ALGORITHM FOR CUSTOMER CLASSIFICATION BASED ON DEMOGRAPHIC DATA AND SHOPPING BEHAVIOR

The increasingly competitive business environment requires companies to understand customer characteristics more deeply so that marketing strategies can be designed more effectively and targeted appropriately. The use of demographic data and purchasing behavior can support customer classification based on their responses to promotions. This study discusses the application of the Random Forest algorithm for customer classification and evaluates the model's performance in distinguishing between customers who respond to promotions and those who do not. The dataset used in this study is Customer Personality Analysis, obtained from Kaggle, which contains demographic variables such as age, income, education level, marital status, and number of children, as well as purchasing behavior variables such as product purchase amount, transaction frequency, and promotional response. Class imbalance was handled using SMOTENC because the dataset consists of both numerical and categorical features. The results show that the Random Forest model with SMOTENC achieved an accuracy of 0.863267, precision of 0.539683, sensitivity of 0.612613, and F1-score of 0.573840. This study is expected to provide an overview of the use of the Random Forest algorithm in customer classification and support the development of more targeted marketing strategies.

Keywords: *Demographic Data, Customer Classification, Purchasing Behavior, Random Forest, SMOTENC.*

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
PENGESAHAN TIM PENGUJI SKRIPSI	iv
HALAMAN PERNYATAAN KEASLIAN	iv
MOTTO	v
PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
ABSTRAK	xiii
ABSTRACT	xiv
I PENDAHULUAN	1
1.1. Rumusan Masalah	9
1.2. Tujuan Penelitian	9
1.3. Manfaat Penelitian	9
1.3.1. Manfaat Teoritis	10
1.3.2. Manfaat Praktis	10
1.4. Batasan Masalah	10
1.5. Sistematika Penulisan	11
II TINJAUAN PUSTAKA	13
2.1. Segmentasi Pelanggan	13
2.2. Data Demografi dan Perilaku Belanja Pelanggan	15
2.3. <i>Preprocessing</i>	17
2.4. SMOTENC	20
2.5. <i>K-Fold Cross Validation</i>	22
2.6. <i>Random Forest</i>	24
2.7. <i>Confusion Matrix</i>	27
2.8. Integrasi Keislaman	29
2.8.1. Klasifikasi dan Optimasi dalam Islam	29
2.8.2. Perilaku Manusia dalam Islam	32

III METODE PENELITIAN	36
3.1. Jenis Penelitian	36
3.2. Sumber Data	36
3.3. Variabel Penelitian	36
3.4. Tahapan Penelitian	41
IV HASIL DAN PEMBAHASAN	45
4.1. Deskripsi Data	45
4.1.1. Analisis Deskriptif	45
4.2. <i>Preprocessing Data</i>	47
4.2.1. Penanganan <i>Missing Value</i>	47
4.2.2. Transformasi Data	47
4.2.3. <i>Encoding Data</i>	48
4.3. Pembagian Data Menggunakan <i>K-Fold Cross Validation</i>	49
4.4. Penanganan Ketidakseimbangan Data	51
4.5. Penerapan Algoritma Random Forest	54
4.5.1. Inisialisasi Parameter	54
4.5.2. Proses Pembentukan Pohon Keputusan pada <i>Random Forest</i>	55
4.5.3. <i>Majority Voting</i>	64
4.6. Hasil Evaluasi Model	66
4.7. Perbandingan Evaluasi Hasil Klasifikasi <i>Random Forest</i> Tanpa SMOTENC dan Dengan SMOTENC	72
4.8. <i>Feature Importance</i>	74
4.9. Integrasi Keislaman	77
V PENUTUP	81
5.1. Kesimpulan	81
5.2. Saran	82
DAFTAR PUSTAKA	82
LAMPIRAN	89

DAFTAR PUSTAKA

- Aburbeian, A. H. M. and Ashqar, H. I. (2023). Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced Data. In *International Conference on Advances in Computing Research*, pages 605–616, Cham. Springer Nature Switzerland.
- Adeniran, I. A., Efunniyi, C. P., Osundare, O. S., and Abhulimen, A. O. (2024). Transforming Marketing Strategies with Data Analytics: A Study on Customer Behavior and Personalization. *International Journal of Scholarly Research in Engineering and Technology*, 4(1):41–51.
- Agustyaningrum, C. I., Gata, W., Nurfalah, R., Radiyah, U., and Maulidah, M. (2020). Komparasi Algoritma Naive Bayes, Random Forest, dan SVM untuk Memprediksi Niat Pembelanja Online. *Jurnal Informatika*, 20(2):164–173.
- Ahmad, A. K., Jafar, A., and Aljoumaa, K. (2019). Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform. *Journal of Big Data*, 6(1):28.
- Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., and Childe, S. J. (2020). How to Improve Firm Performance Using Big Data Analytics Capability and Business Strategy Alignment? *International Journal of Production Economics*, 182:113–131.
- Albalawi, T. and Dardouri, S. (2025). Enhancing Credit Card Fraud Detection Using Traditional and Deep Learning Models with Class Imbalance Mitigation. *Frontiers in Artificial Intelligence*, 8:1643292.
- Alie, J. and Gustriansyah, R. (2024). Customer Segmentation for Digital Marketing Based on Shopping Patterns. *Jurnal Aplikasi Bisnis dan Manajemen*, 10(1):209–216.
- Aribisala, O. E. (2024). Data Quality Assessment and Preprocessing Techniques for Enhancing Machine Learning Model Performance. *IJIRMPS – International Journal of Innovative Research in Multidisciplinary Professional Studies*, 12(3):1–11.

- Ashner, M. C. and Garcia, T. P. (2024). Understanding the Implications of a Complete Case Analysis for Regression Models with a Right-Censored Covariate. *The American Statistician*, 78(3):335–344.
- Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) (2023). Laporan Survei Pengguna Internet Indonesia 2023. <https://survei.apjii.or.id/survei>. Diakses pada 26 Mei 2026.
- Assyarif, E. C. Z. and Nuryana, I. K. D. (2025). Penerapan Klasifikasi Pelanggan Berdasarkan Segmentasi Pelanggan pada UMKM Monex Toys Bekasi. *Modem: Jurnal Informatika dan Sains Teknologi*, 3(3):47–65.
- Bhagat, M. and Bakariya, B. (2025). A Comprehensive Review of Cross-Validation Techniques in Machine Learning. *International Journal on Science and Technology*, 16(1):1–4.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.
- Brijith, A. (2023). Data Preprocessing for Machine Learning. *Cyber Security Insights Magazine*, 3:1–4.
- Dong, S. (2024). Application Research of Multi-Dimensional Customer Behavior Analysis Model in Precision Marketing. *Open Journal of Applied Sciences*, 14(12):3589–3600.
- El-Hajj, M. and Pavlova, M. (2024). Predictive Modeling of Customer Response to Marketing Campaigns. *Electronics*, 13(19):3953.
- Fauzan, R., Nurhayati, S., and Nuraeni, I. (2023). Evaluasi Kinerja Model Klasifikasi Menggunakan Confusion Matrix dan ROC Curve. *Jurnal Teknologi Informasi dan Komputer*, 9(2):115–123.
- Genuer, R., Poggi, J.-M., and Tuleau-Malot, C. (2020). Variable selection using random forests. *Pattern Recognition Letters*, 131:83–91.
- GhorbanTanhaei, H., Boozary, P., Sheykhan, S., Rabiee, M., Rahmani, F., and Hosseini, I. (2024). Predictive Analytics in Customer Behavior: Anticipating Trends and Preferences. *Results in Control and Optimization*, 17:100462.

- Google, Temasek, Bain & Company (2023). e-conomy sea 2023 report. <https://www.bain.com/global/insights/e-conomy-sea-2023/>.
- Hermawan, A. A., Saraswati, G. W., and Kartikadarma, E. (2023). Metode MICE Support Vector Machine (MICE-SVM) untuk Klasifikasi Performance Mahasiswa Merdeka Belajar Kampus Merdeka. *Jurnal Media Informatika Budidarma*, 7(4):1686–1697.
- Huda, M. N. (2025). Analysis of Market Segmentation Based on Geographic, Demographic, Psychographic, and Behavior in CV Hoeda Kharya Santoesa. *International Journal of Review Management Business and Entrepreneurship*, 5(1):98–109.
- Ibrahim, S. P. and Deepa, M. (2025). A Study on Customer Segmentation and Its Impact on Targeting the Right Customers. *International Journal of Novel Research and Development*, 10(4):c126–c134.
- Ilham, I. (2024). Predicting Plant Growth Stages Using Random Forest Classifier: A Machine Learning Approach. *Indonesian Journal of Data and Science*, 5(2):155–165.
- Irnanan, F. D., Hidayah, I., and Nugroho, L. E. (2021). Metode Imputasi pada Data Debit Daerah Aliran Sungai Opak, Provinsi DI Yogyakarta. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 10(4):301–310.
- Kaggle (2023). Customer Personality Analysis & Segmentation. <https://www.kaggle.com/code/hasibalmuzdadid/customer-personality-analysis-segmentation>. Diakses pada 13 Mei 2026.
- Kahraman, A. (2025). Machine Learning Techniques for Improved Prediction of Cardiovascular Diseases Using Integrated Healthcare Data. *Frontiers in Artificial Intelligence*, 8:1694450.
- Khan, A. A., Chaudhari, O., and Chandra, R. (2024). A Review of Ensemble Learning and Data Augmentation Models for Class Imbalanced Problems: Combination, Implementation and Evaluation. *Expert Systems with Applications*, 244:122778.

- Kotler, P. and Keller, K. L. (2009). *Marketing Management*. PT Indeks, Jakarta.
- Maharana, K., Mondal, S., and Nemade, B. (2022). A Review: Data Pre-processing and Data Augmentation Techniques. *Global Transitions Proceedings*, 3(1):91–99.
- Matin, I. M. M. (2023). Hyperparameter Tuning Menggunakan GridsearchCV pada Random Forest untuk Deteksi Malware. *MULTINETICS*, 9(1):43–50.
- Mikalef, P., Pappas, I. O., Krogstie, J., and Giannakos, M. (2020). Big Data Analytics Capabilities: A Systematic Literature Review and Research Agenda. *Information Systems and e-Business Management*, 18(3):547–578.
- Mitra, R., McGough, S. F., Chakraborti, T., Holmes, C., Copping, R., Hagenbuch, N., Biedermann, S., Noonan, J., Lehmann, B., Shenvi, A., Doan, X. V., Leslie, D., Bianconi, G., Sanchez-Garcia, R., Davies, A., Mackintosh, M., Andrinopoulou, E.-R., Basiri, A., Harbron, C., and MacArthur, B. D. (2023). Learning from Data with Structured Missingness.
- Monalisa, S., Juniarti, Y., Saputra, E., Muttakin, F., and Ahsyar, T. K. (2023). Customer Segmentation with RFM Models and Demographic Variable Using DBSCAN Algorithm. *TELKOMNIKA Telecommunication Computing Electronics and Control*, 21(4):742–749.
- Mukherjee, M. and Khushi, M. (2021). SMOTE-ENC: A Novel SMOTE-Based Method to Generate Synthetic Data for Nominal and Continuous Features. *Applied System Innovation*, 4(1):18.
- Nanda, N. N. A., Farida, Y., and Utami, W. D. (2025). Implementation of SMOTE to Improve the Performance of Random Forest Classification in Credit Risk Assessment in Banking. *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, 9(2):158–177.
- Prajapati, P. K. and Uraon, D. (2024). Consumer Behaviour and

- Factors Influencing Consumer Behaviour: A Conceptual Study. *EPRA International Journal of Multidisciplinary Research (IJMR)*, 10(7):1–4.
- Quynh, T. D. and Dung, H. T. T. (2021). Prediction of Customer Behavior Using Machine Learning: A Case Study. In *Proceedings of the 2nd International Conference on Human-centered Artificial Intelligence (Computing4Human 2021)*, volume 3026 of *CEUR Workshop Proceedings*, pages 168–175, Da Nang, Vietnam.
- Rahman, M. M. and Islam, M. M. (2022). A Comprehensive Survey on Random Forest: From Theory to Applications. *Artificial Intelligence Review*, 55(8):5999–6050.
- Reelfs, J. H., Bergmann, M., Hohlfeld, O., and Henckell, N. (2021). Understanding & Predicting User Lifetime with Machine Learning in an Anonymous Location-Based Social Network. In *Companion Proceedings of the Web Conference 2021*, pages 324–331. ACM.
- Sakho, A., Malherbe, E., and Scornet, E. (2026). Do We Need Rebalancing Strategies? A Theoretical and Empirical Study Around SMOTE and Its Variants. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, Tangier, Morocco. PMLR.
- Sciascia, I. (2023). From Market Segmentation to Customer Loyalty. *International Journal of Business and Management*, 18(4):192–199.
- Septa, A. A., Al Farizi, A., Khafid, A. N., Prasetyo, D., Romadhon, N. C., and Utomo, F. S. (2024). Diabetes Detection Optimisation with Hyperparameter Tuning in Random Forest Algorithm. *Journal of Informatics and Interactive Technology*, 1(3):165–177.
- Sun, Y., Liu, H., and Gao, Y. (2023). Research on Customer Lifetime Value Based on Machine Learning Algorithms and Customer Relationship Management Analysis Model. *Heliyon*, 9(2):e13384.
- Ullah, A., Mohmand, M. I., Hussain, H., Johar, S., Khan, I., Ahmad, S., Mahmoud, H. A., and Huda, S. (2023). Customer Analysis Using Machine Learning-Based Classification Algorithms for Effective

Segmentation Using Recency, Frequency, Monetary, and Time. *Sensors*, 23(6):3180.

Wahyudi, A., Talkah, A., and Rusandy, D. S. (2024). Beyond Product Attributes: The Role of Segmentation, Targeting, Positioning in Shaping Consumer Satisfaction in the Agri-Food Sector in Kediri Regency, Indonesia. *Open Access Indonesia Journal of Social Sciences*, 7(5):1789–1801.

Wamba, S. F., Gunasekaran, A., Papadopoulos, T., and Ngai, E. W. T. (2020). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 131:389–400.

Yasodha, P. (2025). Data Preprocessing Methods for Machine Learning: An Empirical Comparison. *International Journal for Multidisciplinary Research*, 7(3):1–7.

Zhang, C. and Ma, Y. (2021). *Ensemble Machine Learning: Methods and Applications*. Springer, New York.

Zhang, D.-y., Lou, H., Liu, J., and Li, B. (2025). Random Forest-Based Identification and Ranking of Predictive Factors for Physical Activity in Chinese College Students. *Frontiers in Public Health*, 13:1699606.

UIN SUNAN AMPEL
S U R A B A Y A