

PREDIKSI CUSTOMER LIFETIME VALUE DENGAN MENGUNAKAN RANDOM FOREST REGRESSOR

SKRIPSI

OLEH:

SRI DWI AFRIANI SEMBIRING

228160041



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2026**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 1/7/26

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah

3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

PREDIKSI CUSTOMER LIFETIME VALUE DENGAN MENGUNAKAN RANDOM FOREST REGRESSOR

SKRIPSI

**Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana di Fakultas Teknik
Universitas Medan Area**

Oleh:

SRI DWI AFRIANI SEMBIRING

228160041

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2026**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 1/7/26

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah

3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Access From (repository.uma.ac.id) 1/7/26

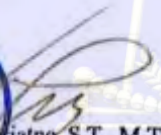
HALAMAN PENGESAHAN

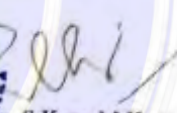
Judul Skripsi : PREDIKSI CUSTOMER LIFETIME VALUE DENGAN
MENGUNAKAN RANDOM FOREST REGRESSOR

Nama : SRI DWI AFRIANI SEMBIRING
NPM : 228160041
Fakultas : Teknik

Disetujui Oleh
Komisi Pembimbing


Andre Hasudungan Lubis, S. Ti, M.Sc
Pembimbing I


Dr. Fery Sutrisno, S.T., M.T.
Dekan Fakultas Teknik


Dr. Fery Sutrisno, S.Kom, M.Kom
Ka. Prodi

Tanggal Lulus: 26 Februari 2026

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.



**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Sri Dwi Afriani Sembiring
NPM : 228160041
Program Studi : Informatika
Fakultas : Teknik
Jenis Karya : Skripsi

Demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul :

**PREDIKSI CUSTOMER LIFETIME VALUE DENGAN MENGGUNAKAN
RANDOM FOREST REGRESSOR**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-eksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan
Pada tanggal : 26 Februari 2026
Yang menyatakan

Sri Dwi Afriani Sembiring
228160041

RIWAYAT HIDUP

Penulis bernama Sri Dwi Afriani Sembiring lahir di Kota Pematangsiantar pada tanggal 17 Juni 2002. Penulis merupakan anak ke-2 dari 2 bersaudara, anak dari Bapak Rafirawan Sembiring dan Ibu Wagiati. Penulis menyelesaikan Pendidikan Sekolah Dasar di SD Negeri 094155 Rambung Merah, Kota Pematangsiantar dan lulus pada tahun 2015, kemudian melanjutkan ke jenjang Sekolah Menengah Pertama di SMP Negeri 2 Kota Pematangsiantar dan lulus pada tahun 2018. Setelah itu penulis melanjutkan pendidikan Sekolah Menengah Atas di SMA Swasta Mars dan lulus pada tahun 2021. Pada tahun 2022 penulis diterima sebagai mahasiswi di Program Studi Informatika, Fakultas Teknik, Universitas Medan Area. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana S. Kom.



KATA PENGANTAR

Assalamu'alaikum warahmatullahi wabarakatuh. Puji Syukur penulisan panjatkan kehadiran Allah SWT atas segala Rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik.

Skripsi ini berjudul **"PREDIKSI CUSTOMER LIFETIME VALUE DENGAN MENGGUNAKAN RANDOM FOREST REGRESSOR"**. Skripsi ini dilanjutkan untuk memenuhi tugas akhir sebagai salah satu syarat kelulusan pada Fakultas Teknik, Program studi Teknik Informatika, Universitas Medan Area.

Dalam proses penyelesaian skripsi ini, penulis menyadari bahwa masih terdapat banyak kekurangan, baik dari segi isi, penulisan, maupun Bahasa. Oleh karena itu pada kesempatan kali ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Dadan Ramdan, M.Eng., M.Sc., selaku Rektor Universitas Medan Area.
2. Bapak Dr. Eng. Supriatno, ST., M.T., selaku Dekan Fakultas Teknik Universitas Medan Area.
3. Bapak Andre Hasudungan Lubis, S. Ti, M. Sc., selaku Dosen Pembimbing yang telah memberikan saya ilmu, arahan, dan serta masukan yang bermanfaat selama proses penyelesaian skripsi ini. Tanpa dukungan dan arahan yang telah beliau berikan kepada penulis selama ini, pasti akan mengalami masa kesulitan dalam proses penyelesaiannya skripsi ini.
4. Bapak Rizki Muliono, S.Kom., M.Kom., selaku Ketua Program Studi Teknik Informatika yang telah memberikan dukungan dan arahan selama masa perkuliahan di Program Studi Teknik Informatika Universitas Medan Area.
5. Teristimewa kepada ayah dan ibu penulis tercinta, yang dengan penuh kasih sayang telah mendidik, mendoakan, serta memberikan dukungan dan motivasi dalam menyelesaikan skripsi ini. Penulis percaya bahwa setiap doa-doa mereka lah yang senantiasa selalu menyelamatkan dan menuntun penulis dalam masa-masa tersulit. Tak lupa kepada abang tercinta yang selalu mendorong saya untuk selalu semangat dan tidak menyerah dalam masa terpuruk selama pengerjaan skripsi.
6. Seluruh dosen dan staf Program Studi Teknik Informatika Universitas Medan Area yang telah memberikan ilmu, bimbingan, serta bantuan selama proses perkuliahan.

7. Seluruh teman-teman Sturnbuk 2022 Teknik Informatika yang telah berbagi pengalaman dan kebersamaan selama masa perkuliahan.
8. Terakhir kepada diri saya sendiri yang senantiasa selalu berjuang, bertahan, dan tidak mengenal kata menyerah walaupun mengalami banyak cobaan dan tantangan. Terimakasih sudah menguatkan diri sendiri selama proses perkuliahan ini.

Penulis berharap semoga Allah SWT senantiasa mengaruniakan berkat dan rahmat-Nya. Semoga skripsi ini dapat memberikan manfaat bagi kita semua.

Medan, 26 Februari 2026
Penulis



Sri Dwi Afriani Sembiring
228160041



ABSTRAK

Customer Lifetime Value (CLV) merupakan salah satu indikator penting dalam bisnis yang digunakan untuk mengukur nilai keuntungan jangka panjang yang dihasilkan oleh seorang pelanggan. Prediksi CLV diperlukan agar perusahaan dapat mengidentifikasi pelanggan yang paling bernilai serta menyusun strategi pemasaran dan pengambilan keputusan bisnis secara lebih efektif. Namun, karakteristik data pelanggan yang kompleks dan memiliki hubungan *non-linear* menyebabkan metode konvensional kurang optimal dalam memprediksi nilai CLV. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma *Random Forest Regressor* (RFR) dalam memprediksi nilai CLV. Data penelitian terdiri dari tiga fitur utama, yaitu *purchase_history*, *tenure*, dan *total_spent*, dengan variabel target berupa nilai CLV. Data dibagi menjadi data latih dan data uji untuk membangun serta mengevaluasi model prediksi. Kinerja model sendiri dievaluasi menggunakan metrik *Mean Square Error* (MSE), *R-squared* (R^2), dan *Mean Absolute Error* (MAE), untuk mengukur tingkat kesalahan prediksi dan kemampuan model dalam menjelaskan variasi data. Pada hasil penelitian menunjukkan model RFR sangat akurat dengan nilai R^2 sebesar 0,9973 yang artinya model mampu menjelaskan sebesar 99,73% variasi data CLV. Namun, tingkat kesalahan prediksi dapat dilihat dari nilai MSE sebesar 348.525 dan MAE sebesar 354,46. Hasil menjelaskan bahwa algoritma RFR terbukti efektif dan akurat dalam memprediksi CLV dan dapat digunakan untuk pengambilan keputusan bisnis sebagai strategi perusahaan.

Kata Kunci: *Customer Lifetime Value*, *Random Forest Regressor*, Prediksi

ABSTRACT

Customer Lifetime Value (CLV) is a crucial business indicator used to measure the long-term profit generated by a single customer. Predicting CLV is essential because it helps companies identify their most valuable customers, allowing them to create more effective marketing strategies and better business decisions. However, customer data is often complex and has non-linear relationships, which makes conventional methods less effective at predicting CLV accurately. Because of this, this study aims to apply the Random Forest Regressor (RFR) algorithm to get better results in predicting CLV. The research data uses three main features: purchase_history, tenure, and total_spent, with CLV as the target variable. The data was split into training and testing sets to build and evaluate the prediction model. To see how well the model performs, it was evaluated using three main metrics: Mean Square Error (MSE), R-squared (R^2), and Mean Absolute Error (MAE). These metrics help measure the prediction error and the model's ability to explain the data's variation. The results show that the RFR model is highly accurate, with an R^2 value of 0.9973, which means the model can explain 99.73% of the variation in the CLV data. Even though the MSE is 348,525 and the MAE is 354.46, these error rates are actually very low when compared to the overall scale of the predicted CLV values. In conclusion, the RFR algorithm proves to be an effective and reliable tool for predicting CLV. It can be used as a helpful resource to support a company's strategic business decision-making.

Keywords: *Customer Lifetime Value, Random Forest Regressor, Prediction*

DAFTAR ISI

HALAMAN PENGESAHAN	i
HALAMAN PERNYATAAN	ii
RIWAYAT HIDUP	iv
KATA PENGANTAR	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii

BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4

BAB II TINJAUAN PUSTAKA	5
2.1 Customer Lifetime Value (CLV)	5
2.2 Regresi.....	5
2.3 Evaluasi Model Regresi	7
2.4 Random Forest Regressor (RFR).....	9
2.5 Penelitian Terdahulu	10

BAB III METODOLOGI PENELITIAN	12
3.1 Tahapan Penelitian	12
3.2 Pengumpulan Data	13
3.3 Prapemrosesan.....	13
3.3.1 Normalisasi Data	14
3.4 Implementasi	15
3.5 Evaluasi	15
3.6 Perhitungan Manual	16

BAB IV HASIL DAN PEMBAHASAN	19
4.1 Karakteristik Data	19
4.2 Korelasi Fitur	21
4.3 Implementasi RFR	23

4.4	Pembahasan.....	26
BAB V KESIMPULAN DAN SARAN		28
5.1	Kesimpulan	28
5.2	Saran.....	29
DAFTAR PUSTAKA.....		30
LAMPIRAN.....		37



DAFTAR TABEL

	Halaman
Tabel 2.1 Penelitian Terdahulu	10
Tabel 3.1 Atribut dari Dataset	13
Tabel 3.2 Rangkuman Dataset	14
Tabel 4.1 Visualisasi Karakteristik Data Menggunakan Violin Plot	19
Tabel 4.2 Statistik Deskriptif Variabel Penelitian.....	21
Tabel 4.3 Hasil Evaluasi Model RFR.....	23
Tabel 4.4 Rangkuman Perbandingan Nilai Aktual dan Prediksi CLV	24



DAFTAR GAMBAR

	Halaman
Gambar 3.1 Tahapan Penelitian.....	12
Gambar 4.1 Correlation Heatmap	22
Gambar 4.2 Grafik Persentase Kesalahan	25
Gambar 4.3 Scatter Plot dari nilai CLV aktual dengan CLV prediksi	26



BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada era perkembangan teknologi yang semakin pesat, dunia bisnis juga mengalami tingkat persaingan yang semakin ketat dan mengalami perubahan dengan cepat. Dalam kondisi ini, peran pelanggan menjadi sangat penting bagi keberlangsungan sebuah perusahaan. Pelanggan merupakan orang yang melakukan transaksi berdasarkan pertimbangan yang sesuai dengan kebutuhan dan keinginan mereka. Pelanggan juga merupakan aset yang sangat berharga bagi perusahaan, karena tidak ada satupun perusahaan yang dapat bertahan tanpa adanya dukungan dan kesetiaan dari pelanggan (Shafly dkk., 2023). Oleh karena itu, perusahaan harus menyadari pentingnya serta peran pelanggan dalam menjaga kelangsungan bisnisnya. Salah satu indikator yang digunakan untuk menilai keberhasilan bisnis terhadap pelanggan adalah *Customer Lifetime Value* (CLV) (Megantara dkk., 2023). CLV merupakan angka dari keuntungan yang diperoleh perusahaan dari seorang pelanggan selama masa keanggotaannya atau selama periode waktu tertentu (Raditya dkk., 2025). Menurut (Wardhana, 2024), pengukuran nilai pelanggan membantu perusahaan dalam menemukan pelanggan yang paling berharga dan mendorong pembuatan keputusan strategis untuk pertumbuhan jangka Panjang. Menurut (Gadgil dkk., 2023), analisis CLV berperan penting dalam membantu sebuah perusahaan untuk mengidentifikasi pelanggan yang paling berharga dan memprediksi potensi keuntungan dimasa mendatang. Pada proses jual beli yang dilakukan melalui internet serta pertukaran informasi antara organisasi dan pihak

luar (Cahyono dkk., 2024). Melalui analisis CLV, perusahaan dapat mengetahui nilai jangka panjang pelanggan dan merancang strategi pemasaran yang lebih efektif.

Seiring dengan pentingnya CLV pada bisnis, dibutuhkan suatu metode yang dapat di terapkan guna memprediksi nilai CLV bagi pelanggan. Salah satu pendekatan yang dapat digunakan untuk memprediksinya adalah *machine learning*. *Machine learning* merupakan elemen penting dalam bidang kecerdasan buatan yang digunakan untuk menyelesaikan berbagai permasalahan (Nurhalizah & Ardianto, 2024). Salah satu metode dari *machine learning* adalah regresi, yaitu alat ukur yang digunakan untuk mengukur sebuah variabel (Yusuf dkk., 2024).

Regresi memiliki berbagai algoritma dengan karakteristik yang berbeda-beda, diantaranya *Linear Regressor*, *Decision Tree*, *Support Vector Machine (SVM)*, dan *Random Forest Regressor (RFR)* (Sharma dkk., 2022). RFR merupakan salah satu algoritma *machine learning* yang menggunakan gabungan beberapa pohon keputusan untuk menghasilkan prediksi yang akurat dan stabil. Algoritma ini juga memiliki keunggulan dalam mengurangi resiko *overfitting*, yang dimana menjadi salah satu masalah dalam algoritma untuk mempelajari mesin lainnya (Efendi dkk., 2024). Selain itu, metode ini dikenal efektif dalam mengelola data yang rumit dan mengurangi kemungkinan terjadinya kesalahan dalam prediksi (Egorenkov, 2024). Penelitian sebelumnya juga menunjukkan bahwa algoritma RFR efektif untuk memprediksi nilai pelanggan di berbagai sektor bisnis. Menurut (Mauludiah dkk., 2024) menunjukkan metode ini sangat akurat dalam memprediksi CLV untuk industri *e-commerce* di Inggris dan Brasil. Selain itu, (Joy dkk., 2025) juga menemukan bahwa RFR dapat menangani data pelanggan yang kompleks dan memberikan hasil prediksi yang konsisten.

Dari uraian diatas, dapat disimpulkan bahwa penelitian ini akan menggunakan algoritma RFR dalam memprediksi CLV, yang diharapkan dapat membantu perusahaan dalam memahami nilai pelanggan secara akurat.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang yang telah disampaikan sebelumnya, maka rumusan masalah dalam penelitian ini adalah bagaimana menerapkan algoritma RFR untuk memprediksi CLV pada dataset pelanggan?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut

1. Data yang digunakan diambil dari dataset pelanggan yang berada di *Kaggle*, mencakup informasi transaksi pelanggan untuk menghitung dan memprediksi nilai CLV.
2. Algoritma yang digunakan adalah RFR sebagai metode utama untuk melakukan prediksi.
3. Evaluasi model menggunakan metrik *Mean Squared Error* (MSE), *R-squared* (R^2) dan *Mean Absolute Error* (MAE), untuk mengukur sejauh mana akurasi dan hasil prediksi.
4. Penelitian ini hanya berfokus pada data historis pelanggan, tanpa memperhitungkan faktor eksternal seperti tren pasar atau promosi.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk menerapkan algoritma RFR dalam memprediksi CLV pada dataset pelanggan.
2. Untuk mengevaluasi kinerja algoritma RFR dalam memprediksi nilai CLV menggunakan metrik evaluasi MSE, R^2 dan MAE.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah untuk menambah referensi dan wawasan mengenai penerapan algoritma RFR dalam memprediksi CLV, serta membantu perusahaan dalam memahami nilai pelanggan guna mendukung pengambilan keputusan dan strategi pemasaran yang lebih efektif.

BAB II TINJAUAN PUSTAKA

2.1 Customer Lifetime Value (CLV)

Customer Lifetime Value (CLV) merupakan pendekatan strategis dalam pengelolaan hubungan pelanggan yang menekankan pada perhitungan total nilai ekonomi yang dihasilkan pelanggan selama menjalani hubungan dengan perusahaan (M. F. Hidayat dkk., 2025). Sementara itu, (Lohonauman, 2020) menjelaskan bahwa CLV adalah metode untuk mengukur keuntungan dari pelanggan, menganalisis strategi pemasaran, dan dari CLV juga dapat digunakan untuk mengembangkan cara berpikir dalam menjalankan kegiatan usaha. Berdasarkan kedua definisi tersebut, dapat disimpulkan bahwa CLV menggambarkan total kontribusi ekonomi pelanggan terhadap perusahaan selama masa hubungan bisnis berlangsung.

Pada penelitian ini, nilai CLV tidak dihitung secara manual lagi dikarenakan dataset yang digunakan telah menyediakan nilai CLV. Oleh karena itu, tahapan penelitian berfokus pada proses pengolahan data dan implementasi algoritma untuk memprediksi CLV yang sudah ada dalam dataset.

2.2 Regresi

Regresi adalah sebuah metode yang digunakan untuk mengukur apakah ada atau tidaknya korelasi antar variabel (Yusuf dkk., 2024). Regresi linear adalah metode statistik yang digunakan untuk mengukur dan memodelkan hubungan antar

variabel *dependen* (Y) dan variabel *independen* (X), untuk melakukan sebuah prediksi berdasarkan data yang sudah ada sebelumnya yang menunjukkan hubungan antara variabel yang menjadi penyebab dan variabel akibat (Fatimah dkk., 2024). Menurut (Rusdy, 2022), regresi linear merupakan metode prediksi yang menggunakan garis lurus untuk menggambarkan keterkaitan diantara dua variabel. Berikut adalah persamaan yang digunakan untuk mendapatkan nilai regresi linear dalam menguji keterkaitan antara variabel bebas dan variabel terikat, yaitu:

$$Y = a + bX \quad (2.4)$$

Dengan menggunakan persamaan ini, untuk nilai X merupakan variabel *independent* dan Y merupakan nilai untuk variabel *dependen*. Berikut merupakan rumus untuk menghitung nilai a dan b:

$$a = \frac{(\sum y \times \sum x^2) - (\sum x (x \times y))}{(n \times \sum x^2) - (\sum x^2)} \quad (2.5)$$

$$b = \frac{(n \times \sum x \times y) - (\sum x \sum y)}{(n \times \sum x^2) - (\sum x^2)} \quad (2.6)$$

Keterangan:

n = Jumlah data

$\sum x$ = Jumlah keseluruhan nilai X

$\sum y$ = Jumlah keseluruhan nilai Y

$\sum x y$ = Jumlah hasil kali Y dan X

$\sum x^2$ = Jumlah kuadrat dari X

Adapun regresi linear berganda yang merupakan perkembangan dari regresi linear sederhana, dengan menggunakan dua atau lebih variabel $x_1, x_2, \dots, x_n$

sebagai variabel independent dan variabel dependen (Khotimah dkk., 2024). Regresi linear berganda dapat mengetahui apakah sebuah variabel independen berpengaruh signifikan terhadap variabel *dependen* (Hariadi dkk., 2024.). Adapun rumus secara sistematis yang dapat dilihat seperti di bawah ini:

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (2.7)$$

Keterangan:

Y = Variabel tak bebas

a = konstanta

$b_1, b_2, \dots, b_n$ = nilai koefisien regresi

$x_1, x_2, \dots, x_n$ = nilai variabel bebas

Sementara itu, model regresi sering digunakan untuk memprediksi CLV karena dapat mengantisipasi hubungan antara atribut pelanggan dan nilai yang akan mendatang bagi perusahaan. Melalui metode regresi, perusahaan dapat memahami faktor-faktor yang berpengaruh terhadap nilai pelanggan dan memperkirakan potensi keuntungan di masa depan (Fahim, 2023).

2.3 Evaluasi Model Regresi

Untuk menilai seberapa baik hasil prediksi dari model regresi diperlukan metrik evaluasi yang berfungsi untuk mengukur sejauh mana prediksi yang dihasilkan oleh model mendekati nilai sebenarnya (Plevris dkk., 2022). Beberapa metrik umum yang digunakan untuk mengevaluasi kinerja model regresi adalah *Mean Squared Error* (MSE), *R-squared* (R^2) dan *Mean Absolute Error* (MAE).

MSE merupakan salah satu metrik yang paling umum digunakan untuk menilai keakuratan prediksi dengan menghitung rata-rata dari selisih kuadrat antara

nilai aktual dan nilai prediksi (Hodson dkk., 2021). Rumus MSE ditunjukkan sebagai berikut:

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n} \quad (2.8)$$

Keterangan:

y_i = Nilai sebenarnya

$\hat{y}_i$ = Nilai prediksi

n = Jumlah sampel data

Nilai MSE selalu bernilai positif dengan rentang antara 0 hingga tak terhingga.

Semakin kecil nilai MSE, semakin baik model dalam melakukan prediksi.

Sementara itu, R^2 adalah nilai yang dapat menjelaskan seberapa besar (persentase) hubungan yang dapat dijelaskan oleh variable bebas terhadap variable terkait (Handayani dkk., 2023). Oleh karena itu, nilai R^2 dapat digunakan untuk menilai seberapa baik model dalam memprediksi data secara tepat serta menilai kecocokan model dengan data yang digunakan (Chicco dkk., 2021). Berikut merupakan rumus yang bisa digunakan:

$$R^2 = \frac{SSR}{SST} \quad (2.9)$$

Rumus bisa menjadi seperti berikut:

$$R^2 = 1 - \frac{\sum (Y - \bar{Y})^2}{\sum (Y - \hat{Y})^2} \quad (2.10)$$

Keterangan:

SSR (*Sum Square Regression*) = $\sum (Y - \bar{Y})^2$

SST (*Sum Square Total*) = $\sum (Y - \hat{Y})^2$

$\sum (Y - \bar{Y})^2$ = Kuadrat selisih nilai Y riil dengan nilai Y prediksi

$\sum(Y - \hat{Y})^2$ = Kuadrat selisih nilai Y riil dengan nilai Y rata-rata

Sedangkan *Mean Absolute Error* (MAE) adalah salah satu metode yang digunakan untuk mengukur tingkat keakuratan model peramalan. Nilai MAE menunjukkan rata – rata kesalahan (*error*) absolut antara hasil peramalan atau prediksi dengan nilai riil. Berikut adalah rumus yang digunakan untuk menghitung MAE:

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_1 - y_i| \quad (2.11)$$

Keterangan:

f_1 = nilai hasil dari peramalan

y_i = nilai sebenarnya

2.4 Random Forest Regressor (RFR)

Random Forest merupakan sebuah model untuk mengklasifikasi dan regresi non-linier yang bergantung pada ansambel pohon keputusan. Dalam *random forest*, terdapat *subset* acak dari sebuah fitur yang dipilih untuk setiap pohon, dan pemisahan fitur terbaik dari sebuah *subset* tersebut yang digunakan untuk membagi setiap pohon (History, 2025). *Random Forest Regressor* (RFR) adalah metode yang menggunakan konsep pohon keputusan dalam bentuk *ensemble*, dimana banyak pohon keputusan dibangun untuk memberikan hasil prediksi yang lebih akurat, serta mengurangi kemungkinan terjadinya *overfitting* yang dapat terjadi pada

decision tree (Taihuttu dkk., 2024). Menurut (R. Hidayat dkk., 2025), RFR merupakan salah satu jenis model regresi yang berlandaskan pada teknik *supervised learning* menggunakan data historis. Model ini merupakan jenis *ensamble learning*, yang berarti menggabungkan beberapa model untuk mendapatkan prediksi yang lebih akurat (Phishing dkk., 2024). Berikut adalah rumus yang digunakan secara matematis:

$$\hat{Y}(x) = \frac{1}{B} \sum_{b=1}^b T_b(x)$$

(2.12)

Berikut adalah *pseudocode* yang menjelaskan tentang algoritma *random forest regressor*.

Algoritma Random Forest Regressor
Srart Input: Dataset D = {(x _i , y _i)}, jumlah pohon B Output: Model prediksi Random Forest Regressor For b = 1 to B: D _b ← sampel bootstrap dari D T _b ← bangun pohon keputusan dari D _b : - Pilih subset acak fitur F - Tentukan pemisahan terbaik berdasarkan MSE minimum Selesai Prediksi akhir: ŷ = (1/B) Σ T _b ((x)) End

2.5 Penelitian Terdahulu

Tabel 2. 1 Penelitian Terdahulu

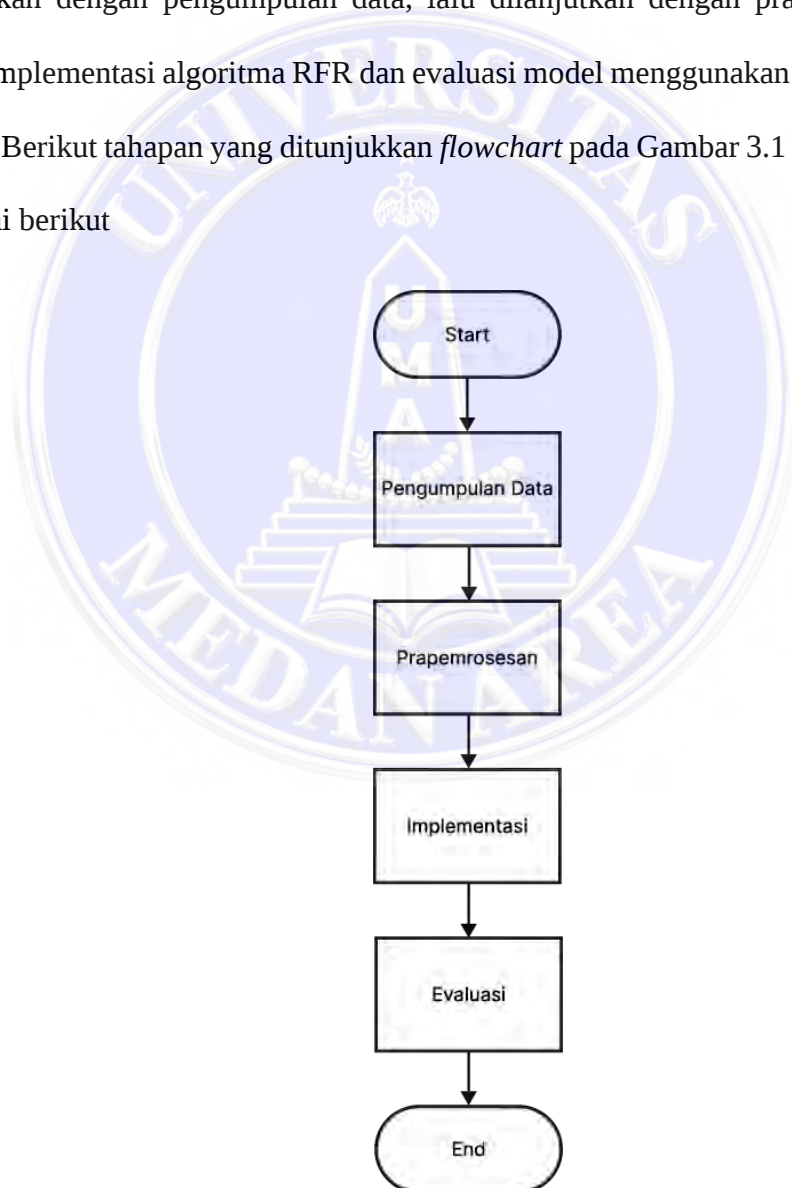
Judul Penelitian, Tahun	Metode	Kesimpulan
“Strategi Penguatan Persepsi Nasabah: pengaruh CRM Terhadap Kualitas Hubungan dan Customer Lifetime Value” (Brando dkk., 2025)	Customer Relationship Management (CRM)	Penelitian ini membuktikan bahwa CRM terbukti berpengaruh positif terhadap relationship quality sebesar 0,57 dan customer lifetime value sebesar 0.19 secara langsung dan 0,44

		secara tidak langsung melalui kualitas hubungan.
“Klasterisasi <i>Customer Lifetime Value</i> Pelanggan Menggunakan K-Means pada Online Travel Agency” (Ananta & Gede, 2024)	<i>K-Means</i>	Penelitian ini membuktikan bahwa kluster 2 memiliki nilai <i>frequency</i> tertinggi yang menunjukkan tingkat loyalitas pelanggan paling tinggi, sedangkan kluster 1 memiliki nilai <i>customer lifetime value</i> terendah sebesar 0,0214.
“Prediksi Jumlah Penjualan Tahun 2024 Menggunakan Metode Regresi Linear” (Studi et al., 2024)	<i>Regresi Linear</i>	Penelitian ini menunjukkan bahwa dengan menggunakan metode regresi linear yang diimplementasikan menggunakan Python mampu memprediksi volume penjualan pada tahun 2024 sebanyak 11.614 pesanan. Evaluasi model menunjukkan performa yang sangat baik dengan nilai R^2 sebesar 0.91 dan tingkat kesalahan yang rendah dengan nilai MAE sebesar 22.89, MSE sebesar 9.13, dan RMSE sebesar 3.02. Penelitian menyimpulkan bahwa penggunaan analisis prediktif dapat membantu UMKM di <i>marketplace</i> dalam mengoptimalkan keuntungan dan mengurangi resiko kerugian.
“Analisis Kinerja Model <i>Regresi</i> dan <i>Machine Learning</i> Untuk Memprediksi Harga Beras” (Hidayat et al., ensemble dalam hal akurasi prediksi dengan stabilitas yang tinggi di berbagai segmen pelanggan 2025)	<i>Regresi</i> dan <i>Machine Learning</i>	Pada penelitian ini menjelaskan bahwa algoritma <i>Linear Regression</i> dalam memproyeksikan hasil panen padi memberikan performa prediksi yang cukup baik dengan menggunakan model yang digunakan menghasilkan nilai MAE sebesar 6550,241810 dan MSE sebesar 121247657,9756, melalui preprocessing data optimal, model mampu memberikan hasil prediksi yang akurat sehingga dapat disimpulkan penggunaan metode regresi dapat menjadi solusi dalam memproyeksikan ketahanan pan di masa depan
“Implementasi Algoritma <i>Random Forest Regression</i> Untuk Memprediksi Penjualan Produk Di Supermarket” ((Hidayat et al., 2025)	<i>Random Forest Regression</i>	Pada penelitian ini menjelaskan bahwa algoritma <i>Random Forest Regression</i> mampu menghasilkan prediksi penjualan supermarket dengan tingkat akurasi yang sangat tinggi yang dapat dilihat dari nilai <i>R-Squared</i> dan <i>OOB Score</i> mendekati 1 dan nilai error yang rendah. Sehingga model efektif untuk digunakan.

BAB III METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Pada penelitian ini, tahapan penelitian dilakukan secara sistematis untuk memperoleh hasil prediksi CLV menggunakan algoritma RFR. Tahap awal yang dilakukan dengan pengumpulan data, lalu dilanjutkan dengan pra-pemrosesan data, implementasi algoritma RFR dan evaluasi model menggunakan MSE, R^2 , dan MAE. Berikut tahapan yang ditunjukkan *flowchart* pada Gambar 3.1 dengan urutan sebagai berikut



Gambar 3. 1 Tahapan Penelitian

3.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini menggunakan data sekunder yang bersumber dari Kaggle dengan judul dataset “*Customer Lifetime Value Prediction Dataset*”. Data tersebut diambil dengan menggunakan tautan sebagai berikut: <https://www.kaggle.com/datasets/sahideseker/customer-lifetime-value-prediction-dataset?resource=download>. Dataset ini berisi sebuah informasi pelanggan yang melakukan sebuah aktivitas transaksi dan hubungan pelanggan dengan sebuah perusahaan.

Data sekunder merupakan data yang diperoleh secara tidak langsung, data ini bisa didapatkan melalui bukti, catatan, buku, jurnal, laporan historis, atau *repository* yang sudah tersusun dalam arsip atau dokumentasi (Arvyanda dkk., 2023). Dataset yang digunakan terdapat 5 atribut yang menggambarkan informasi transaksi pelanggan yang terlampir dalam Tabel 3.1

Tabel 3.1 Atribut dari Dataset

Atribut	Penggunaan
Customer_ID	×
Purchase_history	✓
Tenure	✓
Total_spent	✓
CLV	✓

3.3 Prapemrosesan

Pra-pemrosesan data adalah tahapan yang dilakukan untuk memastikan data yang digunakan dalam penelitian ini bersih, dan layak digunakan dalam proses analisis serta pemodelan (Alnur dkk., 2023). Proses ini penting untuk menghindari data yang tidak *valid*, duplikat maupun data yang tidak konsisten yang dapat mempengaruhi hasil dari prediksi. Kemudian, dilakukan juga proses pengecekan

terhadap nilai yang kosong (*missing values*) dan data yang terduplikasi (*duplicate values*) untuk memastikan konsistensi dataset.

Adapun rangkuman dari data yang akan digunakan pada penelitian ini dapat dilihat pada tabel 3.2

Tabel 3. 2 Rangkuman Dataset

Purchase_history	Tenure	Total_Spent	CLV
39	37	1111.72	3427.8
29	36	8556.68	25670.04
15	24	5317.54	10635.08
...	...	...	...
40	17	8277.79	11726.87

3.3.1 Normalisasi Data

Normalisasi data merupakan sebuah proses yang dilakukan untuk membuat beberapa variabel memiliki rentang nilai yang sama, sehingga tidak ada atribut yang terlalu besar atau terlalu kecil⁴. Proses ini penting agar skala antar atribut dapat sebanding dan tidak ada variabel yang mendominasi proses pembelajaran model (Kusnaldi dkk., 2022). Normalisasi umumnya digunakan pada model yang sensitif terhadap perbedaan skala data. Adapun rumus yang digunakan untuk melakukan normalisasi data yaitu sebagai berikut:

$$\hat{X} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.1)$$

Keterangan:

$\hat{X}$ = Nilai hasil normalisasi

X = Nilai asli

X_{min} = Nilai minimum

X_{max} = Nilai maksimum

3.4 Implementasi

Pada tahap implementasi, model RFR digunakan untuk memprediksi nilai CLV. Nilai CLV yang digunakan dalam penelitian ini merupakan nilai asli yang sudah ada pada dataset, sehingga CLV dijadikan sebagai target value dalam proses pelatihan model. Sebelum pelatihan dilakukan, seluruh atribut pra-pemrosesan digunakan sebagai fitur input. Model kemudian mempelajari hubungan antara fitur tersebut dengan nilai CLV berdasarkan konsep regresi. Proses pelatihan dilakukan menggunakan data latih (*training data*), dan data uji (*testing data*) yang digunakan untuk mengevaluasi kemampuan model dalam melakukan prediksi terhadap data baru. *Random Forest* membangun banyak pohon keputusan (*decision trees*) dan nilai prediksi akhir diperoleh dari rata-rata hasil setiap pohon hingga menghasilkan hasil prediksi yang stabil dan akurat.

3.5 Evaluasi

Tahap evaluasi dilakukan untuk menilai sejauh mana model RFR mampu memprediksi nilai CLV dengan menggunakan evaluasi dua metrik utama, yaitu MSE, R^2 dan MAE. MSE digunakan untuk mengukur rata-rata selisih kuadrat antara nilai prediksi dan nilai aktual. Semakin kecil nilai MSE, maka semakin baik modelnya. Sementara itu, R^2 digunakan untuk mengukur persentase variasi data yang dapat dijelaskan oleh model. Jika mendekati 1 berarti model sangat baik sedangkan jika nilai mendekati 0 maka model kurang baik. Sedangkan pada MAE, nilai yang dihasilkan menunjukkan rata-rata selisih *absolut* antara nilai CLV aktual dengan nilai hasil prediksi. Semakin kecil nilai MAE mendekati angka 0, maka semakin akurat model dalam memprediksi nilai.

3.6 Perhitungan Manual

Pada penelitian ini, dilakukan sebuah perhitungan manual untuk pemahaman yang mendalam terhadap model yang digunakan. Berikut adalah contoh ilustrasi untuk menunjukkan proses perhitungan RFR dengan menggunakan rumus seperti dibawah ini:

$$\hat{Y}(x) = \frac{1}{B} \sum_{b=1}^b T_b(x)$$

Dimana:

B = Jumlah pohon (Trees) dalam ensemble

$T_b(x)$ = Prediksi pohon ke-b untuk sampel x

$\hat{Y}(x)$ = Prediksi akhir model untuk sampel x

Pada implementasi, B misalnya bernil 100. Selanjutnya melakukan perhitungan dengan memiliki 3 sampel dan setiap sampel diprediksi oleh 3 pohon.

Sampel	$T_1(x)$	$T_2(x)$	$T_3(x)$
1	325.000	326.500	324.800
2	1.419.000	1.420.000	1.421.000
3	240.000	240.5000	240.500

Menghitung rata-rata per sampel (prediksi ensemble $\hat{Y}$):

1. Sampel 1:

$$\hat{Y}_1 = \frac{325.000 + 326.500 + 324.800}{3} = 325.433.3333$$

2. Sampel 2:

$$\hat{Y}_2 = \frac{1.419.000 + 1.420.000 + 1.421.000}{3} = 1.420.000.0000$$

3. Sampel 3:

$$\hat{Y}_3 = \frac{240.000 + 240.500 + 240.500}{3} = 240.333.3333$$

Setiap pohon menghasilkan satu nilai prediksi, kemudian RFR menggabungkan hasil dengan cara mengambil nilai rata-rata untuk mengurangi variansi dibandingkan menggunakan satu pohon keputusan saja.

Setelah memperoleh nilai prediksi akhir dari RFR, selanjutnya nilai tersebut digunakan sebagai dasar dalam proses evaluasi model. Oleh karena itu, dilakukan contoh proses nilai actual (Y) dan nilai prediksi ($\hat{Y}$) bukan merupakan hasil model yang sesungguhnya, namun hanya digunakan untuk menjelaskan langkah perhitungan yang akan dilakukan pada tahap selanjutnya. Berikut adalah contoh data yang digunakan:

Data	Y	$\hat{Y}$
1	327.954.5571	325.433.3333
2	1.437.052.9236	1.420.000.000
3	241.938.6594	240.333.333

1. Hitung Error

$$Error = Y - \hat{Y}$$

Dengan menggunakan rumus diatas, berikut adalah hasil dari setiap data

- $327.954.5571 - 325.433.333 = 2.521,2238$
- $1.437.052.9236 - 1.420.000.000 = 17.052,9136$
- $241.938.6594 - 240.333.333 = 1.605,3261$

2. Kuadrat Error

Dari hasil perhitungan error diatas, selanjutnya melakukan perhitungan untuk kuadrat error $(Y_i - \hat{Y}_i)^2$

- $2.521,2238^2 \approx 6.356.569,28$
- $17.052,9136^2 \approx 290.801.862,25$
- $1.605,3261^2 \approx 2.577.071,78$

3. Untuk SSE $\approx 299.735.503,31$

4. Menghitung MSE

Berikut adalah rumus yang digunakan untuk menghitung MSE

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

Untuk $n = 3$

$$MSE = \frac{299.735.503,31}{3} = 99.911.834,44$$

5. Menghitung R^2

a. Pertama hitung rata-rata $\bar{Y}$

$$\bar{Y} \approx 668.982,0434$$

b. Hitung SST

$$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$SST = 888.598.659.880,05$$

c. Hitung R^2

$$R^2 = 1 - \frac{SSE}{SST}$$

$$R^2 = 1 - \frac{299.735.593,31}{888.598.659.880,05} = 0.9996627$$

$$R^2 \approx 0,9996627$$

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini dilakukan dengan tujuan untuk penerapan algoritma RFR dalam memprediksi CLV dan mengevaluasi kinerja model dengan menggunakan metrik MSE dan R^2 , dan MAE. Berdasarkan hasil implementasi yang dilakukan, algoritma RFR berhasil diterapkan dengan hasil nilai MSE sebesar 348525.50 dan MAE sebesar 354.46. Pada nilai MSE yang terlihat besar disebabkan oleh karakteristik nilai CLV antar pelanggan yang sangat bervariasi sehingga selisih nilai yang dikuadratkan menghasilkan angka yang besar. Namun, nilai MAE dengan rata-rata kesalahan prediksi model sangat kecil dibandingkan dengan skala nilai CLV yang diprediksi. Sementara itu, nilai R^2 sebesar 99.73% yang menandakan bahwa model memiliki kemampuan yang sangat baik dalam mempelajari nilai variabel input dan target. Hasil ini diperkuat oleh visualisasi perbandingan nilai aktual dan prediksi CLV dimana sebagian besar titik data berada dekat dengan garis diagonal yang menjelaskan kesesuaian antara hasil prediksi dan data aktual.

Algoritma RFR terbukti mampu memberikan performa yang sangat baik dalam memprediksi nilai CLV. Meskipun terdapat beberapa kesalahan prediksi, namun secara keseluruhan model yang dibangun menunjukkan kinerja yang baik dan dapat digunakan sebagai alat bantu bagi perusahaan dalam memahami nilai pelanggan secara akurat.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dipertimbangkan untuk pengembangan penelitian selanjutnya sebagai berikut.

1. Penelitian selanjutnya disarankan untuk melakukan perbandingan kinerja algoritma RFR dengan model lainnya, seperti Linear Regression, Gradient Boosting Regression, Support Vector Regression, maupun algoritma regresi lainnya. Perbandingan bertujuan untuk mengetahui model mana yang memiliki performa paling optimal dalam memprediksi nilai CLV.
2. Kemudian, disarankan untuk menggunakan data primer dengan fitur yang lebih luas seperti perilaku pembelian, serta faktor eksternal lainnya agar mampu meningkatkan kemampuan model dalam menghasilkan prediksi CLV yang lebih akurat lagi.

DAFTAR PUSTAKA

- Alnur, B., Mulyono, Fitri Amillia, & Sutoyo, S. (2023). JITE (Journal of Informatics and Telecommunication Engineering). *Journal of Informatics and Telecommunication Engineering*, 7(1), 102–111.
- Ananta, I. P., & Gede, I. K. (2024). *Klasterisasi Customer Lifetime Value Pelanggan Menggunakan K-Means pada Online Travel Agency*. 12(4).
- Arvyanda, R., Fernandito, E., & Landung, P. (2023). *Analisis Pengaruh Perbedaan Bahasa dalam Komunikasi Antarmahasiswa*. 1.
- Bauer, J., & Jannach, D. (2021). Improved Customer Lifetime Value Prediction with Sequence-To-Sequence Learning and Feature-Based Models. *ACM Transactions on Knowledge Discovery from Data*, 15(5).
<https://doi.org/10.1145/3441444>
- Brando, A., Sadikin, D. S., Andriani, M., & Budhijana, R. B. (1829). *Strategi Penguatan Persepsi Nasabah : Pengaruh Crm Terhadap Kualitas Hubungan Dan Customer Lifetime Value Di Pt . Bank Mandiri (Persero), Tbk Area Jakarta Tebet Supomo*. 9865, 167–178.
- Cahyono, F. T., Athori, A., & Fauziyah, F. (2024). Penggunaan Activity Based Costing Dan Customer Lifetime Values Guna Pengambilan Keputusan Strategik Profitabilitas Pelanggan. *Jurnal Ilmiah Cendekia Akuntansi*, 8(3), 84. <https://doi.org/10.32503/cendekiaakuntansi.v8i3.4149>

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, 1–24. <https://doi.org/10.7717/PEERJ-CS.623>

Efendi, M. S., Sarwido, & Zyen, A. K. (2024). Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk. *Resolusi : Rekayasa Teknik Informatika dan Informasi*, 5(1), 12–20. <https://doi.org/10.30865/resolusi.v5i1.2149>

Egorenkov, D. (2024). AI-Powered Predictive Customer Lifetime Value: Maximizing Long-Term Profits. *International Journal of Scientific Research and Management (IJSRM)*, 12(09), 7339–7354. <https://doi.org/10.18535/ijssrm/v12i09.em02>

Fahim, S. F. (2023). Customer Lifetime Value Prediction Using the Regression Model. *International Research Journal of Modernization in Engineering Technology and Science*, 07, 3445–3452. <https://doi.org/10.56726/irjmets43604>

Fatimah, F., Casmat, M., Matematika, P. S., & Terbuka, U. (2024). *Pemodelan Regresi Linear Untuk Memprediksi Nilai Penjualan Di Pt Goodiebag Custom*. 5(September), 1–13.

Gadgil, K., Gill, S. S., & Abdelmoniem, A. M. (2023). A meta-learning based stacked regression approach for customer lifetime value prediction. *Journal of Economy and Technology*, 1, 197–207. <https://doi.org/10.1016/j.ject.2023.09.001>

- Handayani, V. A., Reza, W., & Mehmood, S. (2023). Finding the Best Model in Nonlinear Regression: Using the Coefficient of Determination. *Jurnal Varian*, 7(1), 67–76. <https://doi.org/10.30812/varian.v7i1.2322>
- Hariadi, W., Putra, E. D., & Anas, A. (t.t.). *Analisis Regresi Linier Berganda Untuk Memodelkan Faktor Yang Mempengaruhi Nilai Penambahan Utang Tahunan Negara Indonesia*. 10, 36–46.
- Hidayat, M. F., Kalsum, U., Jalu, B., Arjuna, P., Haq, A. N., Tinggi, S., Ekonomi, I., Artikel, I., Value, C. L., Management, C. R., Pelanggan, L., & Pemasaran, S. (2025). *SRAP*. 1, 49–55.
- Hidayat, R., Saputra, H. T., Husnah, M., Bintang, M., Nazhmi, M. N., Azra, J., & Rana, A. (2025). *Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produk Di Supermarket*. 10(1), 101–109.
- History, A. (2025). *Regression 1*. 6(9), 1–14.
- Hodson, T. O., Over, T. M., & Foks, S. S. (2021). Mean Squared Error, Deconstructed. *Journal of Advances in Modeling Earth Systems*, 13(12), 1–10. <https://doi.org/10.1029/2021MS002681>
- Joy, A., Joy, A., & Joy, A. (t.t.). *Predicting Customer Lifetime Value : A Comprehensive Approach with Machine Learning and Deep Learning Models National College of Ireland Supervisor : Jaswinder Singh*.
- Kabasakal, İ. (2020). Customer Segmentation Based On Recency Frequency Monetary Model: A Case Study in E-Retailing. *Bilişim Teknolojileri Dergisi*, 13(1), 47–56. <https://doi.org/10.17671/gazibtd.570866>

Khotimah, A. K., Rahman, A. A., Alam, M. Z., Nur, Y. H., & Aufo, T. R. (2024). *Analisis Regresi Linier Berganda Dalam Estimasi Indeks Pembangunan Manusia di Indonesia Multiple Linear Regression Analysis In Estimating The Human Development Index In Indonesia*. 15(November), 90–99. <https://doi.org/10.30872/eksponensial.v15i2.1318>

Kusnaldi, M. R., Gulo, T., & Aripin, S. (2022). Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal. *Journal of Computer System and Informatics (JoSYC)*, 3(4), 330–338. <https://doi.org/10.47065/josyc.v3i4.2112>

Laksono, B. C., & Wulansari, I. Y. (t.t.). *Pemodelan Dan Penerapan Metode Rfm Pada Estimasi Nilai Konsumen (Customer Lifetime Value) Menggunakan K-Means Clustering (Modeling and Application of the RFM Method on the Estimation of Customer Lifetime)*. 1277–1285.

Laksono, B. C., & Wulansari, I. Y. (2022). Estimating Customer Lifetime Value in the E-Commerce Industry Using Multivariate Analysis. *Proceedings of The International Conference on Data Science and Official Statistics, 2021*(1), 507–518. <https://doi.org/10.34123/icdsos.v2021i1.161>

Lohonauman, H. (2020). Ipteks Perhitungan Costumer Lifetime Value. *Jurnal Ipteks Akuntansi Bagi Masyarakat*, 4(1), 19. <https://doi.org/10.32400/jiam.4.1.2020.29236>

Mauludiah, S. F., Arif, Y. M., Faisal, M., & Putra, D. D. (2024). Struggling Models: An Analysis of Logistic Regression and Random Forest in Predicting Repeat

- Buyers with Imbalanced Performance Metrics. *Applied Information System and Management (AISM)*, 7(2), 31–38.
<https://doi.org/10.15408/aism.v7i2.39326>
- Megantara, R. A., Alzami, F., Akrom, A., Pramunendar, R. A., Prabowo, D. P., Wibowo, S., & Ritzkal, R. (2023). RFM Analysis for Customer Lifetime Value with PARETO/NBD Model in Online Retail Dataset. *Moneter: Jurnal Keuangan dan Perbankan*, 11(2), 284–290.
<https://doi.org/10.32832/moneter.v11i2.409>
- Nurhalizah, R. S., & Ardianto, R. (2024). *Analisis Supervised dan Unsupervised Learning pada Machine Learning : Systematic Literature Review*. 4(1), 61–72.
- Phishing, D., Kamila, A. R., Adikara, F., & Herdian, C. (2024). Analisa Pengaruh Penambahan Fitur dengan Perbandingan Algoritma berbasis Bagging dan. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 10(3), 460–467.
- Plevris, V., Solorzano, G., Bakas, N. P., & Ben Seghier, M. E. A. (2022). Investigation of Performance Metrics in Regression Analysis and Machine Learning-Based Prediction Models. *World Congress in Computational Mechanics and ECCOMAS Congress*, June.
<https://doi.org/10.23967/eccomas.2022.155>
- Raditya, D. D., Adhirajasa, U., & Sanjaya, R. (t.t.). *Analisis Customer Relationship Management Terhadap Customer Life-Time Value Dengan Customer Statisfaction Sebagai Variabel Mediasi*. 5(4), 976–987.

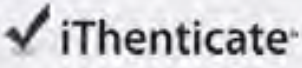
- Rizkyanto, H., & Gaol, F. L. (2023). Customer Segmentation of Personal Credit using Recency, Frequency, Monetary (RFM) and K-means on Financial Industry. *International Journal of Advanced Computer Science and Applications*, 14(4), 152–162. <https://doi.org/10.14569/IJACSA.2023.0140417>
- Rusdy, A. M. A. (2022). Penerapan Metode Regresi Linear pada Prediksi Penawaran dan Permintaan Obat Studi Kasus Aplikasi Point of Sales. 3(2), 121–126.
- Shafly, M. F., Sawiji, H., & Susantiningrum, S. (2023). Analisis kepuasan pelanggan di PT. Komunikasi Selaras Mandiri (CSM). *JIKAP (Jurnal Informasi dan Komunikasi Administrasi Perkantoran)*, 7(3), 184. <https://doi.org/10.20961/jikap.v7i3.60780>
- Sharma, A., Patel, N., & Gupta, R. (2022). Enhancing Customer Lifetime Value Prediction Using Random Forests and Neural Network Ensemble Methods. *European Advanced AI Journal*, 1–24.
- Stays, S., Here, T., Science, S., & Publishing, S. S. (t.t.). *Customer Value Segmentation and Profiling with*. 10(5), 89–96.
- Taihuttu, H. Y., Sitanggang, I. S., & Syaufina, L. (2024). Soil Moisture Prediction Model in Peatland Using Random Forest Regressor. *Barekeng*, 18(4), 2505–2516. <https://doi.org/10.30598/barekengvol18iss4pp2505-2516>
- Wardhana, A. (2024). Mengukur Nilai Pelanggan. Dalam *Customer Value In The Digital Edge – Edisi Indonesia* (Vol. 1, Nomor September).

Yusuf, M. A., Trisnawati, H., Abraham, A., & Rukmana, H. (2024). *Analisis Regresi Linier Sederhana dan Berganda Beserta Penerapannya*. 06(02), 13331–13344.



LAMPIRAN




Similarity Report ID: oId:29477:137733665

PAPER NAME

SRI DWI AFRIANI SEMBIRING_Prediksi Customer Lifetime Value Dengan Menggunakan Random Forest Regressor_sZLqUI2laQlhXM7EAAu4ZjRQ9y4TJTebOVNtMawg.pdf

AUTHOR

SRI DWI AFRIANI SEMBIRING

WORD COUNT

7659 Words

CHARACTER COUNT

49490 Characters

PAGE COUNT

48 Pages

FILE SIZE

1.2MB

SUBMISSION DATE

May 5, 2026 1:58 PM GMT+7

REPORT DATE

May 5, 2026 1:59 PM GMT+7

● **14% Overall Similarity**

The combined total of all matches, including overlapping sources, for each database:

- 13% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 12% Submitted Works database

● **Excluded from Similarity Report**

- Bibliographic material
- Cited material
- Abstract
- Small Matches (Less than 15 words)



UNIVERSITAS MEDAN AREA

FAKULTAS TEKNIK

Kampus 1 | Jalan Kiriwun Nomor 1 Medan Beduga 20131
Kampus 2 | Jalan Semabur Nomor 2 Medan Sei Serapi Nomor 20 A 20132
Website: www.stmik.uma.ac.id E-mail: umk.medan@uma.ac.id

Nomor : 353/FT.001.10/XI/2023

Lamp : -

Jl a l : Penelitian Dan Pengambilan Data Tugas Akhir

27 November 2023

Yth. Wakil Rektor Bidang Misi Sumber Daya Manusia dan Perekonomian
Jln. Kolam No. 1
Di
Medan

Dengan hormat, kami mohon kesediaan bapak kiranya untuk memberikan izin dan kesetiaan kepada mahasiswa kami tersebut dibawah ini :

NO	NAMA	NPM	PRODI
1	Sri Dwi Afriani Sembiring	313160041	Teknik Informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir di Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah sesuai-mat. Untuk tujuan Ilmiah dan Skripsi, yang merupakan salah satu syarat bagi mahasiswa tersebut untuk menyelesaikan studinya pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan tujuan

Prediksi Customer Lifetime Value Dengan Menggunakan Random Forest Regressor.

Mohon kiranya tanggal Surat Izin Pengambilan Data Tugas Akhir agar disesuaikan dengan tanggal Terbitnya Surat ini.

Atas perhatian dan kerja sama yang baik diucapkan terima kasih


 Dekan
 Fakultas Teknik Universitas Medan Area, ST, MT

Tembusan :
 1. Ka. BPMP
 2. Mahasiswa
 3. File



UNIVERSITAS MEDAN AREA

Kampus I : Jalan Kolam Nomor 1 Medan Estate ☎ (061) 7390168, Medan 20223
Kampus II : Jalan Seiabadi Nomor 73 B / Jalan Sei Gerayu Nomor 70 A ☎ (061) 42402994, Medan 20122
Website: www.uma.ac.id E-Mail: uma_medanarea@uma.ac.id

SURAT KETERANGAN SELESAI PENELITIAN

Nomor : 195/UMA/B/01.7/II/2026

Yang bertanda tangan di bawah ini :

Nama : Dr. Ir. Rahmad Syah, M.Kom, IPM, ASEAN Eng, APEC Eng
Jabatan : Wakil Rektor Bidang Mutu Sumber Daya dan Perekonomian
NIDN : 0105058804

Dengan ini menerangkan bahwa mahasiswa yang namanya tercantum di bawah ini :

Nama : Sri Dewi Afriani Sembiring
NPM : 228160041
Program Studi : Teknik Informatika
Fakultas : Teknik
Status : (Mahasiswa / Dosen / Peneliti)

Telah melaksanakan dan menyelesaikan riset (penelitian) di lingkungan Universitas Medan Area dengan rincian sebagai berikut:

Judul Penelitian : Prediksi *Customer Lifetime Value* dengan Menggunakan *Random Forest Regressor*
Lokasi Penelitian : Laboratorium Komputer Program Studi Teknik Informatika Universitas Medan Area
Hasil Penelitian : Berdasarkan penelitian ini, dapat disimpulkan bahwa Nilai MSE sebesar 348.525,50 yang tinggi disebabkan oleh sensitivitas metrik tersebut terhadap rentang nilai CLV yang luas, di mana kesalahan prediksi pada data bernilai tinggi memberikan kontribusi error yang besar. Namun, hal ini tidak menunjukkan kinerja model yang rendah, melainkan dikompensasi oleh nilai R^2 sebesar 0,9973 yang membuktikan bahwa model RFR mampu menjelaskan 99,73% variasi data dengan sangat baik. Visualisasi aktual vs prediksi memperkuat temuan ini, menunjukkan bahwa sebagian besar titik data mengikuti garis diagonal secara akurat, sementara lonjakan MSE hanya dipicu oleh sedikit deviasi pada nilai-nilai ekstrem.
Waktu Pelaksanaan : November – Desember 2025

Berdasarkan laporan hasil riset dan verifikasi data yang kami terima, yang bersangkutan telah menyelesaikan seluruh rangkaian kegiatan penelitiannya dengan baik.

Demikian surat ini diterbitkan untuk dapat digunakan sepenuhnya.

Medan, 02 Februari 2026
Wakil Rektor Bidang Mutu Sumber
Daya dan Perekonomian,



Dr. Ir. Rahmad Syah, M.Kom,
IPM, ASEAN Eng, APEC Eng

