

**PERBANDINGAN PERFORMA DARI *LOGISTIC REGRESSION* DAN *RANDOM FOREST* DALAM
MEMPREDIKSI AKUISISI PELANGGAN**

SKRIPSI

OLEH:

ELISYA APRILIA

218160016



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2025**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 5/8/25

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)5/8/25

PERBANDINGAN PERFORMA DARI *LOGISTIC REGRESSION* DAN *RANDOM FOREST* DALAM MEMPREDIKSI AKUISISI PELANGGAN

SKRIPSI

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

Oleh:

ELISYA APRILIA

218160016

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2025**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 5/8/25

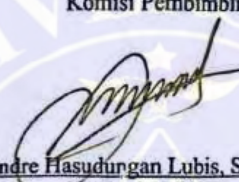
1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Access From (repository.uma.ac.id)5/8/25

HALAMAN PENGESAHAN

Judul Skripsi : Perbandingan Performa Dari *Logistic Regression* Dan *Random Forest* Dalam Memprediksi Akuisisi Pelanggan
Nama : Elisya Aprilia
NPM : 218160016
Fakultas : Teknik

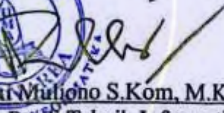
Disetujui Oleh
Komisi Pembimbing


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




Dekan S.T. M.T
Dekan Fakultas Teknik




Kaprodi S.Kom, M.Kom
Kaprodi Teknik Informatika

Tanggal Lulus: 06 Maret 2025

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.

Medan, 06 Maret 2025



Elisya Aprilia
218160016

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

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

Nama : Elisya Aprilia
NPM : 218160016
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis Karya : Skripsi

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

"Perbandingan Performa dari *Logistic Regression* dan *Random Forest* Dalam Memprediksi Akuisisi Pelanggan"

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan 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 : 06 Maret 2025
Yang menyatakan


Elisya Aprilia
218160016

RIWAYAT HIDUP



Elisya Aprilia adalah nama penulis skripsi ini. Lahir di Medan pada tanggal 3 April 2002. Penulis merupakan anak Pertama dari Tiga bersaudara, dari pasangan Bapak Zabidi dan Ibu Kuahate.

Penulis pertama kali masuk pendidikan SD di Negeri 066434 Kotabangun pada tahun 2008 dan tamat pada tahun 2014 pada tahun yang sama penulis melanjutkan pendidikan ke SMP Negeri 43 Medan dan tamat pada tahun 2017. Setelah tamat di SMP, penulis melanjutkan ke SMA Laksamana Martadinata Medan dan tamat pada tahun 2020. Pada tahun 2021, penulis melanjutkan Kuliah dan terdaftar sebagai Mahasiswa di Universitas Medan Area Fakultas Teknik Prodi Teknik Informatika dan tamat pada tahun 2025.

Segala puji bagi Allah memberikan daya kepada penulis serta dukungan, ketekunan dan motivasi tinggi untuk terus belajar dan berusaha. Penulis telah berhasil menyelesaikan pengerjaan tugas akhir skripsi ini. Semoga dengan penulisan skripsi ini mampu memberikan banyak manfaat dan kontribusi yang baik bagi dunia pendidikan.

Akhir kata penulis mengucapkan rasa syukur yang sebesar – besarnya atas terselesainya skripsi ini yang berjudul **”Perbandingan Algoritma Logistic Regression dan Random Forest dalam Memprediksi Akuisisi Pelanggan”**

KATA PENGANTAR

Segala Puji dan Syukur penulis panjatkan kehadirat Allah SWT atas limpahan rahmat, karunia, serta hidayah-Nya, penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Perbandingan Performa Algoritma *Logistic Regression* dan *Random Forest* dalam memprediksi akuisisi pelanggan”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Medan Area.

Penyusunan skripsi ini tentu tidak terlepas dari berbagai tantangan, baik dalam proses pengumpulan data, pengolahan data, hingga analisis hasil. Namun berkat bimbingan, dukungan, doa, dan semangat dari berbagai pihak, penulis dapat menyelesaikan skripsi ini dengan baik. Oleh karena itu, pada kesempatan ini, penulis dengan penuh rasa hormat menyampaikan terimakasih kepada:

1. Rektor Universitas Medan Area, Bapak Prof. Dr. Dadan Ramdan, M.Eng., M.Sc yang telah menyediakan fasilitas dan kesempatan untuk menempuh pendidikan di lingkungan Universitas Medan Area.
2. Dekan Fakultas Teknik Universitas Medan Area, Bapak Dr. Eng., Supriatno, S.T, M.T yang telah memberikan dukungan akademik selama masa studi.
3. Ketua Program Studi Teknik Informatika, Bapak Rizki Muliono S.Kom, M.Kom yang telah memberikan arahan dan motivasi dalam menjalani perkuliahan hingga penyusunan skripsi.
4. Dosen Pembimbing, Bapak Andre Hasudungan Lubis, S.Ti, M.Sc selaku dosen pembimbing penulis yang dengan sabar membimbing, memberikan

arahan, masukan, dan motivasi selama proses penyusunan skripsi ini sehingga skripsi ini dapat terselesaikan dengan baik.

5. Seluruh Dosen di Prodi Teknik Informatika yang telah memberikan ilmu, wawasan, serta inspirasi selama masa perkuliahan.
6. Kedua Orang Tua saya Bapak Zabidi dan Ibu Kuahate, dua orang yang sangat berjasa dalam hidup saya. Terimakasih atas segala pengorbanan dan tulus kasih yang diberikan, kerja keras dan keringat yang bercucuran engkau tukarkan menjadi nafkah demi anakmu bisa sampai pada tahap ini, mereka memang tidak sempat merasakan pendidikan di bangku perkuliahan, namun mereka mampu senantiasa memberikan yang terbaik, tak kenal lelah mendoakan, serta memberikan dukungan hingga penulis mampu menyelesaikan studinya sampai meraih gelar sarjana. Terimakasih atas segala hal yang kalian berikan yang tak terhitung jumlahnya. Semoga bapak dan ibu sehat, panjang umur dan bahagia selalu.
7. Teman seperjuangan penulis, Gurrly. Terimakasih atas semangat dan support yang tidak pernah padam, atas persahabatan yang tulus dan semua pengalaman yang telah memperkaya perjalanan penulis baik suka maupun duka, kegagalan dan keberhasilan yang telah dilalui bersama, kebersamaan yang terjalin, kerjasama dalam mengerjakan tugas dan skripsi, hingga tawa yang menguatkan di tengah tekanan menjadi bagian yang sangat berharga dalam perjalanan ini. Semoga Gurrly senantiasa diberikan kesuksesan dan kebahagiaan dalam setiap langkah yang ditempuh.
8. IT Support, Bang Robby Kurniawan Sari S.T. Terimakasih atas bantuan dalam segala lika – liku dimasa perkuliahan, sangat membantu kelancaran

proses administrasi akademik, serta penyelesaian berbagai kebutuhan dokumen yang diperlukan untuk penyusunan dan penyelesaian skripsi ini.

9. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan dukungan baik secara langsung maupun tidak langsung.

10. Terakhir, Elisya Aprilia, ya! diri saya sendiri. Apresiasi sebesar – besarnya karena telah bertanggung jawab untuk menyelesaikan apa yang telah dimulai. Terimakasih karena terus berusaha dan tidak menyerah serta senantiasa menikmati setiap proses yang tidak mudah. Terimakasih sudah mau bertahan.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, baik dari segi isi, metode, maupun penyajiannya. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi penyempurnaan karya ini kedepannya. Akhir kata, semoga skripsi ini dapat memberikan manfaat dan menjadi tambahan referensi bagi pihak – pihak yang membutuhkan.

Medan, 06 Maret 2025



Elisya Aprilia
218160016

ABSTRAK

Akuisisi pelanggan merupakan langkah awal yang penting dalam upaya mencari dan mempertahankan pelanggan yang tepat, yakni dengan memastikan bahwa pelanggan baru menerima perhatian yang sesuai. Proses ini menjadi sangat penting bagi keberlangsungan dan pertumbuhan bisnis. Dalam pasar yang bersifat kompetitif, kemampuan untuk menarik dan mempertahankan pelanggan baru dapat menjadi faktor penentu kesuksesan bagi bisnis. Penelitian ini bertujuan untuk membandingkan dua algoritma klasifikasi, yaitu *Logistic Regression* dan *Random Forest*, dalam memprediksi akuisisi pelanggan dengan menggunakan dataset sekunder yang diambil dari Kaggle. Metodologi penelitian mencakup beberapa tahap, yakni eksplorasi data untuk memahami karakteristik dataset, prapemrosesan data untuk mempersiapkan informasi yang relevan, serta evaluasi kinerja model menggunakan *Confusion Matrix*. Penelitian ini menggunakan metrik presisi, recall, dan f1-score untuk mengukur performa kedua algoritma. Hasil penelitian menunjukkan bahwa Algoritma *Random Forest* memiliki akurasi yang lebih tinggi dibandingkan dengan *Logistic Regression*, dengan nilai akurasi 95%, presisi 95%, recall 95% dan f1-score 95%. Di sisi lain, algoritma *Logistic Regression* memiliki nilai akurasi sebesar 85%, presisi 85%, recall 85% dan f1-score 85%. Temuan ini menunjukkan bahwa *Random Forest* lebih efektif dalam memprediksi akuisisi pelanggan. Penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam pemilihan model *Machine Learning* yang optimal untuk memprediksi akuisisi pelanggan, sehingga bisnis dapat merancang strategi pemasaran yang lebih tepat dan efisien, meningkatkan kontribusi terhadap pengembangan bisnis, dan meningkatkan nilai pelanggan jangka panjang.

Kata kunci: Akuisisi pelanggan, *Logistic Regression*, *Random Forest*, *Machine Learning*, Prediksi Pelanggan.

ABSTRACT

Customer acquisition is an important first step in the effort to find and retain the right customers, by ensuring that new customers received appropriate attention. This process became very important for the sustainability and growth of a business. In a competitive market, the ability to attract and retain new customers could be a determining factor in business success. This research aimed to compare two classification algorithms, namely Logistic Regression and Random Forest, in predicting customer acquisition using a secondary dataset taken from Kaggle. The research methodology included several stages, namely data exploration to understand dataset characteristics, data preprocessing to prepare relevant information, and model performance evaluation using the Confusion Matrix. This research used precision, recall, and f1-score metrics to measure the performance of the two algorithms. The results of the research showed that the Random Forest algorithm had higher accuracy compared to Logistic Regression, with an accuracy value of 95%, precision 95%, recall 95%, and f1-score 95%. On the other hand, the Logistic Regression algorithm had an accuracy value of 85%, precision 85%, recall 85%, and f1-score 85%. These findings indicated that Random Forest was more effective in predicting customer acquisition. This research was expected to provide significant contributions in selecting the optimal Machine Learning model for predicting customer acquisition, so that businesses could design more accurate and efficient marketing strategies, improve business development contributions, and increase long-term customer value.

Keywords: Customer Acquisition, Logistic Regression, Random Forest, Machine Learning, Customer Prediction.



DAFTAR ISI

HALAMAN PENGESAHAN.....	i
HALAMAN PERNYATAAN	ii
RIWAYAT HIDUP	iv
KATA PENGANTAR.....	v
ABSTRAK.....	viii
ABSTRACT.....	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan	3
1.4. Manfaat Penelitian	4
1.5. Batasan Masalah	4
BAB II TINJAUAN PUSTAKA.....	6
2.1. Akuisisi Pelanggan	6
2.2. Klasifikasi.....	7
2.3. <i>Logistic Regression</i>	8
2.4. <i>Random Forest</i>	9
2.5. <i>Confusion Matrix</i>	10
2.6. Penelitian Terdahulu	10
BAB III METODOLOGI PENELITIAN.....	12
3.1. Alat dan Bahan	12
3.1.1. Tahapan Penelitian	13
3.1.2. Tinjauan Literatur	13
3.1.3. Pengumpulan Data	14
3.1.4. Prapemrosesan	17
3.1.5. EDA (<i>Exploratory Data Analysis</i>).....	18
3.1.6. Implementasi.....	19
3.1.7. Evaluasi	20
3.2. Perhitungan Manual.....	21
BAB IV HASIL DAN PEMBAHASAN.....	26
4.1. Data Penelitian.....	26
4.2. Prapemrosesan.....	26
4.2.1. Memeriksa Data Null	27
4.2.2. Transformasi Data.....	28
4.2.3. Normalisasi Data.....	29

4.3.	EDA	31
4.3.1.	Statistik Deskriptif	31
4.3.2.	Histogram	31
4.3.3.	Violin Plot	32
4.3.4.	Korelasi	34
4.4.	Implementasi Algoritma dan Evaluasi.....	36
4.5.	Perbandingan dengan Penelitian Terdahulu	38
 BAB V KESIMPULAN DAN SARAN		39
5.1.	Kesimpulan.....	39
5.2.	Saran	40
 DAFTAR PUSTAKA.....		41



DAFTAR TABEL

	Halaman
Tabel 2.1. <i>Confusion Matrix</i>	10
Tabel 2.2. Penelitian Terdahulu	10
Tabel 3.1 Rangkuman Dataset	16
Tabel 3.2 Dataset yang telah di Transformasi.....	16
Tabel 4.1. Rangkuman Data.....	27
Tabel 4.2. Dataset yang telah Ditransformasi	30
Tabel 4.3 Dataset yang telah Dinormalisasi.....	30
Tabel 4.4 Statistik Deskriptif	31
Tabel 4.5 Hasil Performa Klasifikasi Masing – Masing Algoritma.....	37



DAFTAR GAMBAR

	Halaman
Gambar 3.1. Tahapan Penelitian	13
Gambar 4.1. Memeriksa Data Null	28
Gambar 4.2. Jenis Data pada Dataset	29
Gambar 4.3. Histogram Produk	32
Gambar 4.4. Violin Plot.....	33
Gambar 4.5. Tabel Korelasi.....	35
Gambar 4.6. Hasil Tabel Korelasi.....	35
Gambar 4.7. <i>Confusion Matrix Logistic Regression dan Random Forest</i>	36



BAB I PENDAHULUAN

1.1. Latar Belakang

Logistic Regression adalah algoritma klasifikasi *Machine Learning* yang digunakan untuk memprediksi probabilitas variabel tidak bebas secara kategorikal. Dalam *Logistic Regression*, variabel yang terikat adalah variabel biner yang berisi data berkode 1 (ya) atau 0 (tidak). Sehingga menjadikan algoritma *Logistic Regression* mempunyai kemampuan untuk memberikan prediksi yang mudah dipahami terkait hasil biner (Wilson dkk., 2024). Hal ini dibuktikan dengan beberapa penelitian terdahulu yang menggunakan *Logistic Regression* untuk memprediksi berbagai objek.

Selain algoritma *Logistic Regression*, *Machine Learning* juga memiliki algoritma yang cukup baik dalam melakukan klasifikasi, yaitu *Random Forest*. *Random Forest* adalah algoritma klasifikasi yang terdiri dari banyak pohon keputusan (Manzali & Elfar, 2023). Algoritma ini menggunakan *bagging* dan *random feature* saat membangun setiap pohon individu yang mencoba membuat hutan pepohonan yang tidak berkorelasi (Iranzad & Liu, 2024). *Random Forest* merupakan representatif dari metode ansambel dimana selama pembangunan pohon keputusan komponen pada setiap langkah pemilihan terpisah, *Random Forest* memilih secara acak subset fitur terlebih dahulu, kemudian melakukan prosedur pemilihan pembagian secara konvensional dalam subset fitur yang dipilih (Hu & Szymczak, 2023).

Seperti halnya algoritma *Logistic Regression*, algoritma *Random Forest* juga banyak digunakan pada penelitian terdahulu untuk melakukan suatu proses klasifikasi. Pada penelitian yang dilakukan oleh Rahmat (2023), algoritma *Random Forest* memiliki nilai akurasi yang paling tinggi dibandingkan algoritma – algoritma lainnya dalam melakukan prediksi penilaian kepuasan penumpang. Kedua algoritma tersebut sudah banyak diterapkan pada kasus yang berkaitan dengan klasifikasi pelanggan.

Pelanggan adalah seseorang yang akan membeli barang atau produk yang dijual dalam jangka beberapa waktu tertentu (Widjanarko dkk., 2023). Orang yang dapat disebut pelanggan yaitu apabila seseorang memiliki interaksi yang kuat dan melakukan pembelian secara berulang kali, dan jika tidak berulang kali mereka hanya disebut sebagai pembeli (Fahreza & Putra, 2024). Kualitas menjadi tujuan utama bagi pihak penyedia layanan untuk selalu melakukan inovasi dengan tujuan agar konsumen merasa puas atas layanan yang diberikan. Kualitas adalah sesuatu yang diputuskan oleh pelanggan (Mahira dkk., 2021) dimana bahwa kualitas dinilai berdasarkan pengalaman nyata dari pelanggan dan konsumen terhadap produk (Adhari, 2021).

Akuisisi pelanggan diperlukan untuk menemukan pelanggan yang tepat sebelum suatu bisnis dapat memutuskan bagaimana cara mendapatkannya, akuisisi pelanggan penting untuk memastikan bahwa pelanggan baru layak mendapatkan perhatian yang sama yang diberikan kepada pelanggan potensial yang menarik agar dapat mempertahankan pelanggan baru (Gill-Wiehl dkk., 2022).

Menurut Yang dkk. (2020), proses akuisisi pelanggan merupakan suatu hal yang sangat penting dari bagian manajemen pelanggan. Hal ini didasarkan pada

akuisisi pelanggan adalah interaksi pertama yang terjadi antara pelanggan dengan bisnis dan dapat berkemungkinan adanya keberlanjutan hingga pelanggan melakukan penggunaan atas jasa, atau pembelian produk pada transaksi pertama (Siregar dkk., 2024). Sehingga, proses yang dilakukan harus berlangsung secara efisien. Lebih lanjut, hal ini dapat menerapkan teknik *Machine Learning* untuk mendukung bisnis dalam menentukan akuisisi pelanggan (Zadoo dkk., 2022). Namun, penggunaan model yang sesuai cukup penting untuk dipertimbangkan.

Dengan membandingkan kedua algoritma tersebut, bisnis dapat memilih pendekatan yang paling sesuai untuk strategi akuisisi pelanggan. Pemilihan model yang tepat akan memberikan dampak besar bagi bisnis dalam merumuskan strategi yang efektif dan responsif terhadap kebutuhan pelanggan (Alfyando dkk., 2024).

1.2. Rumusan Masalah

Penelitian ini merumuskan masalah sebagai berikut yaitu bagaimana perbandingan performa dari algoritma *Logistic Regression* dan *Random Forest* dalam memprediksi akuisisi pelanggan.

1.3. Tujuan

Penelitian ini bertujuan untuk membandingkan performa dari algoritma *Logistic Regression* dan *Random Forest* dalam memprediksi akuisisi pelanggan.

1.4. Manfaat Penelitian

Adapun manfaat penelitian untuk:

1. Peningkatan Pengalaman Pelanggan: Dengan wawasan yang lebih baik tentang preferensi pelanggan, bisnis dapat menciptakan produk dan layanan yang lebih sesuai, meningkatkan kepuasan dan loyalitas pelanggan.
2. Identifikasi Faktor Kunci: Penelitian membantu bisnis memahami faktor – faktor yang mempengaruhi keputusan pelanggan, sehingga strategi pemasaran dapat dirancang dengan lebih tepat.
3. Identifikasi Risiko: Penelitian dapat membantu mengidentifikasi risiko dalam akuisisi pelanggan, memungkinkan bisnis untuk mengambil langkah lebih awal.

1.5. Batasan Masalah

Adapun rumusan masalah berdasarkan latar belakang diatas:

1. Dalam penelitian ini, berfokus pada perbandingan performa dari dua algoritma *Machine Learning* yaitu, *Logistic Regression* dan *Random Forest* dalam memprediksi akuisisi pelanggan.
2. Data yang digunakan berasal dari data pelanggan yang sudah ada, data sekunder yang diperoleh dari kaggle :

<https://www.kaggle.com/datasets/rabieelkharoua/consumer-electronics-sales-dataset?resource=download>

3. Atribut yang digunakan pada data adalah *KategoriProduk*, *MerkProduk*, *HargaProduk*, *UsiaPelanggan*, *JenisKelamin*, *FrekuensiPembelian*, *KepuasanPelanggan*, *NiatMembeli*.
4. Penelitian ini menggunakan *Confusion Matrix* untuk mengevaluasi performa model.



BAB II TINJAUAN PUSTAKA

2.1. Akuisisi Pelanggan

Akuisisi pelanggan merupakan strategi bisnis untuk mendapatkan laba dengan mencari pelanggan baru. Akuisisi pelanggan bisa dilakukan dengan identifikasi, dan mengembangkan hubungan dengan pelanggan baru (Wahyudi & Sutoyo, 2022). Sangat penting bagi bisnis untuk melakukan strategi akuisisi pelanggan untuk membentuk masa depan bisnis yang berkelanjutan, bisnis tidak akan hidup jika tidak memiliki pelanggan (Martiwi, 2021).

Pelanggan merupakan konsumen yang melakukan pembelian atau pemakaian secara berkala terhadap suatu produk atau jasa, yang dimana konsumen merupakan orang yang membeli atau menggunakan suatu produk atau jasa. Pelanggan adalah individu atau organisasi yang sudah efektif melakukan pembelian (Dachi, 2020). Definisi pelanggan adalah setiap orang yang menuntut pemberian jasa (perusahaan) untuk memenuhi suatu standar kualitas pelayanan tertentu, sehingga dapat memberi pengaruh pada performansi (*performance*) pemberi jasa (perusahaan) tersebut. Dengan kata lain, pelanggan adalah orang – orang atau membeli produk yang tidak tergantung pada suatu produk, tetapi produk yang tergantung pada orang tersebut. Oleh karena ini pembeli atau pengguna suatu produk maka harus diberi kepuasan. Jadikan pelanggan sebagai raja, lakukan yang mereka inginkan, sehingga tetap nyaman dan loyal terhadap bisnis (Dwiastanti dkk., 2022).

2.2. Klasifikasi

Kecerdasan buatan atau *AI (Artificial Intelligence)* adalah suatu sistem yang dikembangkan dalam bidang studi yang dimodelkan pada mesin maupun komputer yang dapat memiliki kecerdasan yang sama atau bahkan bisa lebih seperti manusia yang ditandai dengan kemampuan beradaptasi dan pengambilan keputusan (Manongga dkk., 2022). *AI* tidak menggantikan peran manusia, tetapi peran *AI* sebagai pendukung kinerja SDM, oleh karena itu perlunya pengembangan kompetensi oleh SDM yaitu kompetensi yang tidak dapat dilakukan oleh *AI* salah satunya adalah meningkatkan *soft skill* SDM (Devianto & Dwiasnati, 2020).

Dengan adanya teknologi *Artificial Intelligence* dan *Machine Learning* bisnis dapat berinteraksi dengan konsumen secara otomatis karena teknologi *Artificial Intelligence* dan *Machine Learning* dapat memberikan jawaban yang benar secara otomatis dari konsumen, sudah banyak bisnis yang menggunakan *AI* berupa chatbot yang memungkinkan pelaku usaha berinteraksi dengan pelanggannya 24 jam (Pambudi dkk., 2020).

Machine Learning (pembelajaran mesin) adalah cabang dari Kecerdasan Buatan yang berfokus pada pengembangan algoritma untuk komputer belajar dari data dan membuat keputusan. *Machine Learning* pada bisnis dapat membantu untuk memprediksi status pengiriman barang. Data – data berupa durasi waktu yang direncanakan agar dapat menjadi masukan untuk memprediksi status pengiriman (Hartawan, 2021). Selain itu, dengan adanya *Machine Learning* bisa menampilkan barang apa saja yang sering dibeli dan dapat membantu konsumen lainnya untuk

melihat barang yang banyak diminati dengan mempunyai value yang tinggi (Junaedy & Asfiah, 2024).

2.3. Logistic Regression

Logistic Regression adalah metode statistik yang digunakan untuk memprediksi perilaku pelanggan berdasarkan variabel independent tertentu. *Logistic Regression* digunakan untuk memodelkan probabilitas dari sebuah hasil biner, seperti apakah seseorang pelanggan akan membeli produk, mendaftar layanan, atau melakukan pembelian ulang (Watugilang & Heikal, 2024). Variabel dependen dalam *Logistic Regression* adalah biner (misalnya, “ya” atau “tidak”, “membeli” atau “tidak membeli”), model ini dapat digunakan untuk memprediksi perilaku pelanggan berdasarkan faktor riwayat pembelian, atau interaksi pelanggan dengan bisnis (Wahyudia & Kadyanana, 2023).

Logistic Regression memberikan probabilitas untuk setiap pelanggan, sehingga memungkinkan bisnis untuk menargetkan upaya pemasaran yang lebih efektif kepada pelanggan yang berisiko tinggi *churn* atau pelanggan yang berpotensi tinggi melakukan pembelian ulang (Dewi dkk., 2024). *Logistic Regression* memiliki bentuk dasar dari regresi linier yakni sebagai berikut.

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2.1)$$

$$f(y_i) = \pi \frac{y_i}{i} (1 - \pi \frac{y_i}{i})^{1-y_i} \quad (2.2)$$

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \quad (2.3)$$

$$g(x) = \ln \left[\frac{\pi(x)}{1-\pi(x)} \right] = \beta_0 + \beta_1 x \quad (2.4)$$

$$\iota(\beta) = \prod_{i=1}^n \pi(x_i) y_i^{1-\pi(x_i)} [1-\pi(x_i)]^{1-y_i} \quad (2.5)$$

$$L(\beta) = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1-y_i) \ln[1 - \pi(x_i)]\} \quad (2.6)$$

2.4. Random Forest

Random Forest adalah algoritma *Machine Learning* yang digunakan untuk memprediksi perilaku pelanggan dengan membangun beberapa pohon keputusan (decision tree) (Sotarjua & Santoso, 2022). *Random Forest* dapat membantu bisnis dalam berbagai aspek yang berkaitan dengan analisis perilaku pelanggan, seperti memprediksi *churn* (berhenti berlangganan), segmentasi pelanggan, rekomendasi produk, dan deteksi anomali dalam data pelanggan.

Random Forest mampu menangani berbagai jenis variabel dengan baik, baik numerik maupun kategoris (Ciptady dkk., 2022). Misalnya, jika sebuah bisnis ingin memprediksi apakah seseorang pelanggan akan berhenti menggunakan layanan mereka (*churn*), *Random Forest* dapat menganalisis sejumlah variabel seperti riwayat pembelian, frekuensi interaksi dengan layanan, dan demografi untuk memberikan prediksi yang akurat (Lestari dkk., 2024). *Random Forest* memiliki bentuk dasar dari regresi linier yakni sebagai berikut.

$$\iota(y) = \underset{n=1}{\operatorname{argmax}}_c \left(\sum_{n=1}^N I_{h_n(y)=c} \right) \quad (2.7)$$

$$Gini(D) = 1 - \sum_{i=1}^c p_i^2 \quad (2.8)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2 \quad (2.9)$$

2.5. Confusion Matrix

Confusion Matrix adalah salah satu cara untuk mengukur performa dari model *Machine Learning* yang dibuat. *Confusion Matrix* berisi informasi mengenai nilai aktual dan klasifikasi dari model tersebut dan disajikan dalam bentuk tabel.

Tabel 2.1. Confusion Matrix

PREDICTION / ACTUAL	POSITIVE	NEGATIVE
POSITIVE	True Positive (TP)	False Negative (FN)
NEGATIVE	False Negative (FN)	True Negative (TN)

Penjelasan dari tabel 2.1 yaitu TP adalah nilai benar positif, yang berarti angka prediksi positif sama dengan nilai aktual yang benar. FP adalah nilai salah positif, yang berarti angka prediksi negatif tidak sama dengan nilai aktual yang benar. FN adalah salah negatif, berarti angka prediksi tidak sama dengan nilai aktual yang benar. TN adalah nilai benar negatif, yang berarti angka prediksi negatif sama dengan nilai aktual benar.

2.6. Penelitian Terdahulu

Tabel 2.2. Penelitian Terdahulu

Penulis (tahun)	Judul	Objek	Metode	Hasil
(A. F. Dewi & Pratiwi, 2021)	Analisis Regresi Logistik Biner pada Pengaruh Harga, kualitas pelayanan dan promosi	Pengaruh Harga	Regresi Logistik	Harga yang lebih murah memiliki pengaruh signifikan terhadap

	terhadap kepuasan pelanggan dalam menggunakan jasa layanan grab di kabupaten Lamongan			kepuasan pelanggan. Pelanggan yang merasakan harga lebih murah kemungkinan memiliki 1.475 kali lebih besar untuk puas dibandingkan dengan harga yang mahal.
(Usman dkk., 2023)	Analisis perilaku pelanggan menggunakan Metode <i>Ensemble Logistic Regression</i>	Perilaku pelanggan	<i>Ensemble Logistic Regression, boosting dan bagging</i>	Model <i>Logistic Regression</i> dasar memiliki akurasi sebesar 59%, <i>boosting</i> 76% dan <i>bagging</i> 75%. Hasil ini menunjukkan bahwa teknik ensemble, khususnya <i>boosting</i> memberikan performa yang lebih baik dalam mengklasifikasi preferensi pelanggan terhadap transmisi kendaraan
(Rizkya, 2023)	Penerapan <i>naïve bayes classifier</i> untuk analisis sentimen data ulasan aplikasi <i>e-commerce shopee</i> pada <i>google play store</i>	<i>e-commerce shopee</i>	<i>Naïve Bayes</i>	Sentimen ini dianalisis menggunakan <i>Naïve Bayes Classifier</i> , sebuah algoritma yang cocok untuk klasifikasi teks. Model ini menghasilkan tingkat akurasi sebesar 79% berdasarkan perhitungan <i>confusion matrix</i> , yang menilai performa model dalam memprediksi sentimen dengan benar.

BAB III

METODOLOGI PENELITIAN

3.1. Alat dan Bahan

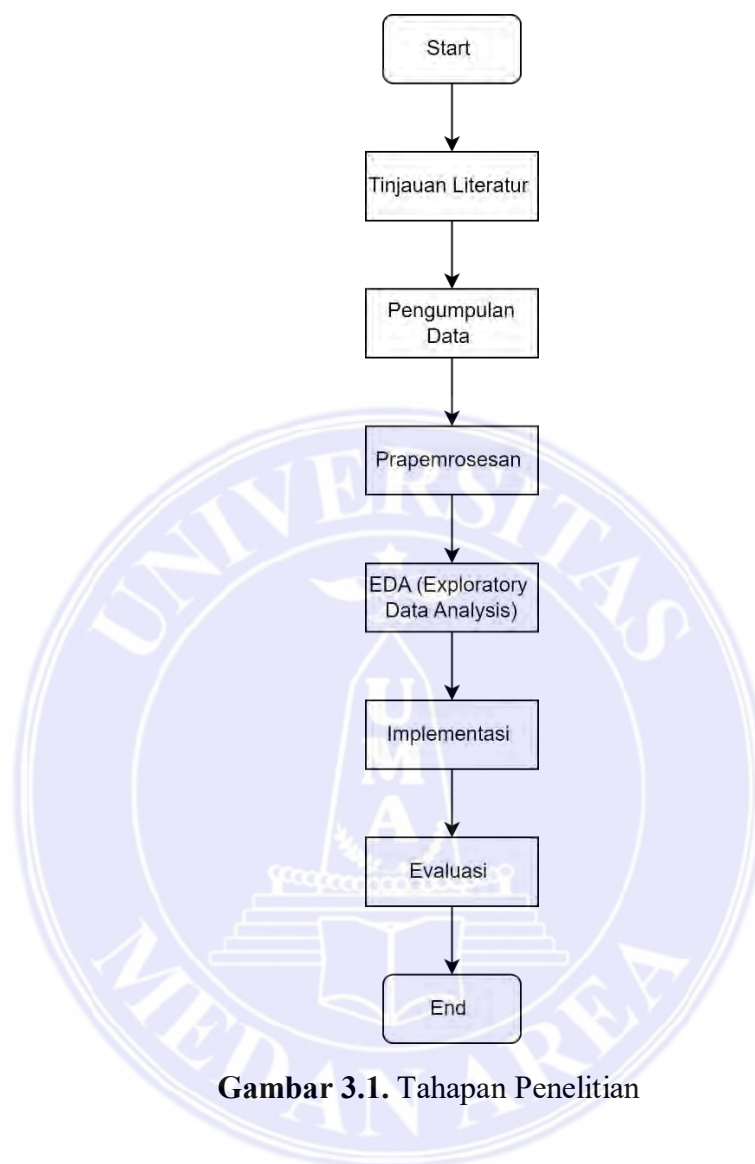
Hardware (perangkat keras) yang dipakai dalam pembuatan skripsi ini adalah sebagai berikut:

1. Laptop / PC
2. *Processor* intel celeron n4020
3. RAM 4 GB

Software (perangkat lunak) yang dipakai dalam pembuatan skripsi ini adalah sebagai berikut:

1. Menggunakan sistem informasi operasi Windows 10
2. Menggunakan *Google Collab*
3. Menggunakan *Microsoft Exel*
4. Menggunakan *Web Browser Google Chrome*

3.1.1. Tahapan Penelitian



Gambar 3.1. Tahapan Penelitian

3.1.2. Tinjauan Literatur

Dalam memprediksi akuisisi pelanggan, *Logistic Regression* dan *Random Forest* sering digunakan sebagai algoritma klasifikasi. ***Logistic Regression*** adalah model linier yang biasa dipakai karena kesederhanaan dan interpretabilitasnya. Model ini cocok untuk memprediksi kejadian biner, seperti apakah pelanggan akan melakukan pembelian atau tidak.

Sebaliknya, **Random Forest** adalah algoritma *ensemble learning* yang membangun banyak pohon keputusan untuk mencapai prediksi yang lebih akurat. Algoritma ini juga unggul dalam menangani data yang lebih kompleks namun interpretabilitasnya lebih rendah.

Perbandingan antara keduanya menemukan bahwa meskipun *Logistic Regression* lebih mudah diinterpretasi, *Random Forest* memberikan hasil yang lebih baik dalam hal akurasi data pada pelanggan yang lebih kompleks. Namun, interpretabilitas *Random Forest* lebih rendah dibandingkan *Logistic Regression*. Dengan demikian, tinjauan ini menunjukkan bahwa *Logistic Regression* maupun *Random Forest* memiliki keunggulan masing – masing dalam memprediksi akuisisi pelanggan.

Penelitian ini akan membandingkan performa kedua algoritma tersebut untuk menemukan model terbaik dalam konteks yang relevan bagi bisnis. Sehingga penelitian ini melakukan tinjauan literatur yang terkait dengan algoritma – algoritma yang digunakan. Selain itu, juga meninjau mengenai teknik penerapannya dan tahapan evaluasi yang dilakukan. Adapun sumber – sumber yang digunakan berasal dari artikel jurnal, buku, prosiding, dan laporan penelitian.

3.1.3. Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan menggunakan data sekunder. Data sekunder adalah sumber data penelitian yang diperoleh secara tidak langsung melalui media perantara misalnya riset-riset kepustakaan yaitu dengan mengumpulkan, membaca dan melalui buku artikel, jurnal, skripsi, atau dari internet yang berkaitan dengan riset ini (Al Ghivary dkk., 2023).

Data ini diperoleh dari berbagai sumber online seperti, website dan platform digital yang menyediakan informasi terkait. Dataset yang digunakan dalam penelitian ini mencakup informasi yang relevan dengan perilaku pelanggan seperti, frekuensi kunjungan, riwayat transaksi dan data demografis seperti usia dan gender. Dataset tersebut terdiri dari beberapa atribut yang mempresentasikan informasi terkait perilaku dan demografi pelanggan. Setiap baris dalam dataset menggambarkan satu pelanggan, sementara kolomnya berisi berbagai fitur untuk analisis prediktif. Atribut pertama adalah **ID Pelanggan** yang merupakan kode unik untuk setiap pelanggan. Selanjutnya, **Jumlah Pembelian** yaitu total transaksi yang dilakukan oleh pelanggan selama periode tertentu. Atribut berikutnya adalah **Status Pelanggan** yang merupakan variabel target dalam penelitian ini. Status ini mengindikasikan apakah pelanggan tersebut diindikasikan sebagai akuisisi pelanggan dengan nilai biner dimana 1 mewakili pelanggan yang aktif dan 0 untuk pelanggan yang tidak terakuisisi.

Selain perilaku atribut demografi seperti **Jenis Kelamin** juga disertakan untuk memberikan konteks tambahan yang relevan dalam analisis yang membantu dalam melihat apakah ada perbedaan pola perilaku antara kelompok pelanggan yang berbeda. Demografi juga diperkuat dengan **Kelompok Umur** yang membagi pelanggan ke dalam rentang usia seperti 18 – 25 tahun, 26 – 35 tahun untuk memahami kecenderungan pembelian berdasarkan usia. Selain itu, **Jumlah Produk yang Dibeli** mencatat total produk yang dibeli dalam satu transaksi atau selama periode waktu tertentu yang memberikan wawasan tentang perilaku pembelian. Adapun rangkuman dari data yang akan digunakan pada penelitian ini dapat dilihat pada Tabel 3.1.

Tabel 3.1 Rangkuman Dataset

ProductID	ProductCategory	ProductBrand	ProductPrice	CustomerAge	CustomerGender	PurchaseFrequency	CustomerSatisfaction	PurchaseIntent
5874	Smartphones	Other Brand	31294	18	0	2	1	0
5875	Smart Watches	Samsung	98038	35	1	7	2	1
5876	Tablets	Samsung	26067	63	0	1	5	1
...	...	...	...	...	...	...	...	...
14873	Smartphones	Sony	23155	62	0	15	2	1

Tabel 3.2 Dataset yang telah di Transformasi

ProductID	ProductCategory	ProductBrand	ProductPrice	CustomerAge	CustomerGender	PurchaseFrequency	CustomerSatisfaction	PurchaseIntent
5874	5	103	31294	18	0	2	1	0
5875	4	104	98038	35	1	7	2	1
5876	6	104	26067	63	0	1	5	1
...					...	...	...	...
14873	5	105	23155	62	0	15	2	1

3.1.4. Prapemrosesan

Prapemrosesan data adalah langkah penting untuk memastikan data yang digunakan sudah bersih dan siap di analisis. Pertama, perlu memeriksa apakah ada data yang hilang atau kosong (null). Misalnya, ada kolom pada dataset yang seharusnya berisi informasi transaksi pelanggan, tapi pada beberapa baris mungkin kosong atau tidak terisi kita harus mengatasi masalah ini dengan menghapus data yang kosong. Selanjutnya, kita juga melakukan pembersihan data. Ini berarti akan mencari dan menghapus data yang tidak wajar atau berulang. Misalnya, jika ada pelanggan yang secara tidak masuk akal mencatat transaksi negatif atau transaksi yang terlalu besar kita perlu menghapusnya agar data tidak bias dan hasil analisis lebih akurat.

Selain itu, pada penelitian ini juga perlu melakukan normalisasi data khususnya data yang terkait dengan transaksi pelanggan. Hal ini penting karena biasanya nilai transaksi pelanggan bisa sangat bervariasi, ada yang belanja sedikit dan ada yang sangat banyak. Normalisasi adalah proses transformasi data mentah menjadi bentuk yang lebih konsisten dan dapat dibandingkan, normalisasi membantu menghindari perbedaan besar dalam rentang variabel yang bisa mempengaruhi hasil analisis atau model prediktif. Dengan normalisasi penelitian ini dapat menyamakan skala semua data, misalnya mengubah semua nilai transaksi ke rentang antara 0 sampai 1. Ini membantu algoritma *Logistic Regression* dan *Random Forest* agar bekerja lebih baik karena semua data memiliki skala yang seimbang. Rumus untuk melakukan normalisasi data dapat dilihat pada Persamaan (3.1).

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

3.1.5. EDA (*Exploratory Data Analysis*)

Tujuan utama dari EDA adalah untuk memahami karakteristik dataset yang digunakan serta mendapatkan insight awal yang dapat berguna saat pemodelan.

Dataset yang digunakan pada penelitian ini terdiri dari variabel yang mempengaruhi kemungkinan akuisisi pelanggan, seperti usia, riwayat pembelian dan beberapa variabel lainnya yang berkaitan dengan akuisisi pelanggan.

Langkah berikutnya dalam EDA adalah memahami distribusi dari variabel target yaitu akuisisi pelanggan. Menggunakan visualisasi, kita dapat melihat apakah data tersebut tidak seimbang atau seimbang. Jika dataset tersebut tidak seimbang, contohnya (jumlah pelanggan yang diakuisisi jauh lebih sedikit daripada yang tidak diakuisisi) kita perlu mempertimbangkan teknik menyeimbangkan seperti oversampling untuk mengatasi masalah dalam tahap pemodelan ini.

Pada tahap akhir EDA, dataset siap untuk di bagi menjadi data latih dan data uji. Proses pembagian ini penting untuk memastikan bahwa model yang dibangun dapat dievaluasi dengan baik pada data yang belum pernah dilihat sebelumnya. Pemisahan data biasanya dilakukan dengan proporsi 70% untuk pelatihan dan 30% untuk pengujian.

3.1.6. Implementasi

Langkah pertama adalah mempersiapkan data yang sudah melalui tahap EDA, kemudian dipecah menjadi data pelatihan dan pengujian untuk memastikan bahwa evaluasi model dilakukan secara valid pada data yang tidak dilihat selama pelatihan.

Logistic Regression adalah metode statistik yang umum digunakan untuk memprediksi variabel target biner (0 atau 1), yang dimaksud variabel target adalah apakah pelanggan diakuisisi (1) atau tidak (0). Setelah membagi data menjadi fitur (X) dan target (Y), kemudian model *Logistic Regression* dilatih menggunakan data pelatihan. Pada tahap ini semua akan diskalakan menggunakan metode *Standard Scaler* untuk memastikan bahwa *Logistic Regression* dapat mengkonvergensi dengan baik karena model ini sensitif terhadap skala data. Evaluasi *Logistic Regression* dilakukan menggunakan *accuracy precision, recall, dan f1-score*. *Confusion Matrix* juga dihasilkan untuk melihat apakah prediksi benar atau salah terutama jika data target tidak seimbang.

Kemudian, model *Random Forest* diterapkan. *Random Forest* adalah algoritma *ensemble* yang terdiri dari beberapa *decision tree*. Algoritma ini cukup kuat untuk menangani data yang kompleks dan memiliki fitur penting seperti menangani variabel yang memiliki korelasi tinggi serta fitur yang bersifat non – linear. Setelah data pelatihan disiapkan, model *Random Forest* dilatih dengan menggunakan parameter default terlebih dahulu untuk mendapatkan kinerja dasar.

Setelah model *Random Forest* selesai dilatih prediksi data uji dilakukan, dan hasilnya dievaluasi menggunakan metrik yang sama seperti *Logistic Regression* yaitu *accuracy*, *precision recall* dan *f1-score*.

Untuk membandingkan kedua model, metrik evaluasi dari *Logistic Regression* dan *Random Forest* akan dibandingkan secara langsung. Pada akhirnya, hasil perbandingan ini akan menunjukkan model mana yang lebih baik dalam memprediksi akuisisi pelanggan. *Logistic Regression*, meskipun sederhana dan lebih mudah diinterpretasikan mungkin memiliki keterbatasan dalam menangani data yang kompleks. Sementara itu, *Random Forest* dengan kemampuan ensemblenya diharapkan dapat menangani lebih baik variasi pada data dan memberikan hasil yang lebih akurat.

3.1.7. Evaluasi

Evaluasi ini dilakukan untuk membandingkan kedua model dengan menggunakan metrik – metrik seperti *accuracy*, *precision*, *recall* dan *f1-score* untuk mengukur efektivitas masing – masing model dalam memprediksi apakah pelanggan akan diakuisisi atau tidak. Adapun rumus yang digunakan untuk mencari nilai – nilai tersebut adalah sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3.2)$$

$$Precision = \frac{Tp}{TP+FP} \quad (3.3)$$

$$Recall = \frac{TP}{TP+FN} \quad (3.4)$$

$$F1 - score = \frac{2TP}{2TP+FP+FN} \quad (3.5)$$

3.2. Perhitungan Manual

Perhitungan manual digunakan untuk menunjukkan pemahaman yang mendalam terhadap konsep atau metode yang diterapkan, serta untuk memverifikasi hasil yang diperoleh dari perangkat lunak. Perhitungan manual juga bisa membantu memperjelas alur logika yang digunakan dalam analisis data, sehingga pembaca dapat mengikuti langkah – langkah perhitungan secara jelas.

Berikut adalah perhitungan manual dengan rumus *Logistic Regression*

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Dimana:

Y = variabel terikat (nilai yang diprediksi)

X = variabel bebas

β_0 = konstanta

β_1 = koefisien regresi (nilai peningkatan ataupun penurunan)

ε = galat acak

- Y = **PurchaseIntent** (target yang diprediksi: 0 atau 1)
- X₁ hingga X₇ adalah variabel input: *ProductCategory*, *Product Brand*, *ProductPrice*, *CustomerAge*, *CustomerGender*, *PurchaseFrequency*, dan *CustomerSatisfaction*.

Langkah – langkah perhitungannya:

1. Bentuk Umum

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Untuk menyederhanakan, kita asumsikan beberapa nilai koefisien sederhana dari pelatihan model dengan data lengkap:

2. Asumsi Koefisien

$$\beta_0 = -5 \text{ (intercept)}$$

$$\beta_1 = 0.1 \text{ untuk } ProductCategory$$

$$\beta_2 = 0.01 \text{ untuk } ProductBrand$$

$$\beta_3 = 0.00001 \text{ untuk } ProductPrice$$

$$\beta_4 = -0.05 \text{ untuk } CustomerAge$$

$$\beta_5 = 0.2 \text{ untuk } CustomerGender$$

$$\beta_6 = 0.3 \text{ untuk } PurchaseFrequency$$

$$\beta_7 = 0.4 \text{ untuk } CustomerSatisfaction$$

3. Hitungan Manual

Dengan data yang ada:

$$ProductCategory = 5$$

$$ProductBrand = 103$$

$$ProductPrice = 31,294$$

$$CustomerAge = 18$$

$$CustomerGender = 0$$

$$PurchaseFrequency = 2$$

$$CustomerSatisfaction = 1$$

Substitusi nilai X_1 hingga X_7 kedalam persamaan:

$$Y = -5 + (0.1 \times 5) + (0.01 \times 103) + (0.00001 \times 31294) + (-0.05 \times 18) + (0.2 \times 0) + (0.4 \times 1)$$

$$Y = -5 + 0.5 + 1.03 + 0.31294 - 0.9 + 0 + 0.6 + 0.4$$

$$Y = -3.05706 + \varepsilon$$

Karena ε adalah error (biasanya diabaikan dalam perhitungan manual sederhana), maka $Y = -3.05706$

4. Interpretasi

Dalam *Regresi Logistic*, hasil ini adalah skor linier yang belum dikonversi menjadi probabilitas.

Kita bisa menghitung probabilitas Y^* dengan mengkonversinya menggunakan fungsi sigmoid:

$$Y^* = \frac{1}{1 + e^{-(-3.05706)}} \approx 0.045$$

Dengan probabilitas 0.045, hasil regresi ini memprediksi bahwa *PurchaseIntent* = 0 (tidak ada niat untuk membeli), karena probabilitasnya jauh lebih rendah dari 0.5

Berikut adalah perhitungan manual dengan rumus **Random Forest**.

$$\hat{y} = \operatorname{argmax}_c \left(\sum_{n=1}^N I_{h_n(y)=c} \right)$$

Penjelasan:

$\hat{y}$: Prediksi akhir dari *Random Forest* atau algoritma *ensemble* lainnya (dalam hal ini adalah *PurchaseIntent*).

c : Kelas yang mungkin (dalam kasus ini $c = 0$ atau $c = 1$ karena targetnya adalah *PurchaseIntent* yang biner)

$h_n(X)$: Prediksi dari *decision tree* ke- n unput input x (input data)

I : fungsi indikator, yang bernilai 1 jika kondisi dalam kurung terpenuhi (jika prediksi tree ke- n adalah c) dan bernilai 0 jika tidak

N : jumlah *decision trees* (dalam contoh kita, $N = 3$)

$\sum$: penjumlahan untuk semua *decision trees*.

Langkah – langkah perhitungan dengan rumus tersebut:

Untuk menghitung $\hat{y}$ menggunakan rumus tersebut, kita akan menghitung jumlah masing – masing kelas (0 dan 1) dari ketiga *decision trees* yang kita bangun sebelumnya voting adalah kelas dengan jumlah prediksi terbanyak.

Data dari 3 trees:

Tree 1: prediksi 0

Tree 2: prediksi 1

Tree 3: prediksi 0

1. Hitung voting untuk kelas $c = 0$

Hitung jumlah *decision trees* yang memprediksi kelas 0

Tree 1: prediksi 0 $\rightarrow I(h_1(x) = 1) = 1$

Tree 2: prediksi 1 $\rightarrow I(h_2(x) = 1) = 0$

Tree 3: prediksi 0 $\rightarrow I(h_3(x) = 1) = 0$

Jumlah untuk kelas 0:

$$\sum_{n=1}^3 I(h_n(x) = 0) = 1 + 0 + 1 = 2$$

2. Hitung *voting* untuk kelas $c = 1$

Hitung jumlah *decision tree* yang memprediksi kelas 1.

Tree 1: prediksi 0 $\rightarrow I(h_1(x) = 1) = 0$

Tree 2: prediksi 1 $\rightarrow I(h_2(x) = 1) = 1$

Tree 3: prediksi 0 $\rightarrow I(h_3(x) = 1) = 0$

jumlah untuk kelas 1:

$$\sum_{n=1}^3 I(h_n(x) = 1) = 0 + 1 + 0 = 1$$

3. Tentukan kelas dengan *voting* terbanyak

Membandingkan hasil *voting* untuk kedua kelas:

Jumlah *voting* untuk kelas 0: 2

Jumlah *voting* untuk kelas 1: 1

kelas dengan *voting* terbanyak adalah **kelas 0**

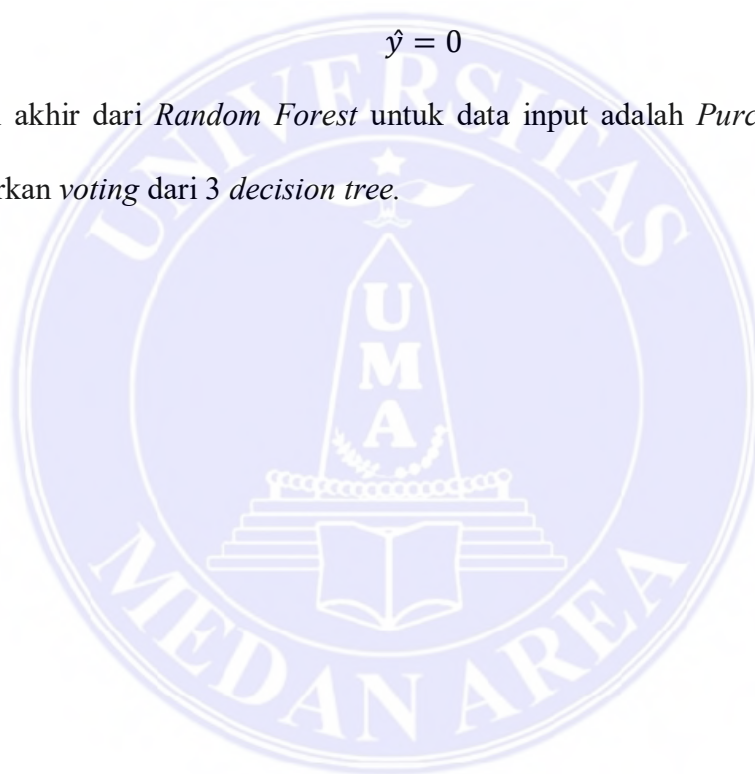
berdasarkan rumus *majority voting* $\hat{y} = \operatorname{argmax} (\sum_{n=1}^3 I(h_n(x) = c))$, prediksi

final $\hat{y}$ untuk data ini adalah:

$$\hat{y} = 0$$

Prediksi akhir dari *Random Forest* untuk data input adalah *PurchaseIntent* = 0

berdasarkan *voting* dari 3 *decision tree*.



BAB V KESIMPULAN DAN SARAN

5.1. Kesimpulan

1. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa dari dua algoritma *Machine Learning* yaitu, *Logistic Regression* dan *Random Forest* dalam memprediksi akuisisi pelanggan. Penilaian kinerja dilakukan berdasarkan 4 metrik evaluasi utama yaitu, akurasi, presisi, recall dan f1-score untuk menentukan algoritma yang paling optimal dalam konteks data pelanggan yang digunakan.
2. Berdasarkan hasil evaluasi menggunakan *Confusion Matrix* algoritma *Logistic Regression* menghasilkan akurasi sebesar 85% (0,85) menunjukkan performa yang baik dalam memprediksi akuisisi pelanggan. Sedangkan, algoritma *Random Forest* menunjukkan performa unggul dengan menghasilkan akurasi sebesar 95% (0.95). Hal ini menunjukkan kemampuan yang sangat baik dalam memprediksi akuisisi pelanggan. Hasil ini menunjukkan bahwa *Random Forest* lebih efektif dalam meminimalkan kesalahan prediksi dibandingkan *Logistic Regression* menjadikannya algoritma yang lebih unggul untuk digunakan dalam strategi akuisisi pelanggan.

5.2. Saran

1. Dalam memprediksi akuisisi pelanggan selain *Logistic Regression* dan *Random Forest*, terdapat berbagai algoritma yang dapat digunakan sebagai alternatif yaitu *K-Nearest Neighbors* (KNN), *Naïve Bayes*, atau jaringan syaraf tiruan untuk mendapatkan variasi performa yang lebih luas.
2. Penelitian ini hanya menggunakan data sekunder, sehingga diharapkan pada penelitian berikutnya dapat menggunakan data primer. Penggunaan data primer dalam memprediksi akuisisi pelanggan dapat memberikan hasil yang lebih relevan karena menggunakan data aktual yang sesuai dengan permasalahan.
3. Pada penelitian ini, faktor – faktor yang digunakan untuk memprediksi akuisisi pelanggan meliputi usia pelanggan, jenis kelamin pelanggan, harga produk, kepuasan pelanggan, dan yang lainnya. Sehingga dapat menambahkan faktor lain seperti promosi dan penawaran khusus, seperti diskon atau *bundling* produk yang juga menjadi faktor penting dalam mendorong keputusan pembeli.
4. Selain *Confusion Matrix* terdapat pula berbagai metode yang dapat digunakan untuk mengevaluasi model prediksi akuisisi pelanggan seperti, *Kappa Statistic (Cohen's Kappa)*, *Kappa Statistic* mengukur akurasi model dengan mempertimbangkan peluang kesepakatan secara acak. Nilai ini sangat cocok pada data yang memiliki distribusi kelas tidak seimbang, karena memberikan evaluasi yang lebih realistis terhadap performa model dibandingkan metrik akurasi biasa.

DAFTAR PUSTAKA

- Adhari, I. Z. (2021). *Kepuasan Pelanggan & Pencapaian Brand Trust* (Vol. 1). CV. Penerbit Qiara Media.
- Al Ghivary, R., Mawar, M., Wulandari, N., Srikandi, N., & others. (2023). Peran Visualisasi Data Untuk Menunjang Analisa Data Kependudukan Di Indonesia. *PENTAHHELIX*, 1(1), 57–62.
- Alfyando, M., Anggraeny, F. T., & Sihananto, A. N. (2024). Perbandingan Algoritma Random Forest dan Logistic Regression Untuk Analisis Sentimen Ulasan Aplikasi Tumbuh Kembang Anak Di Play Store. *Jurnal Sistem Informasi Dan Ilmu Komputer*, 2(1), 77–86.
- Ciptady, K., Harahap, M., Jonvin, J., Ndruru, Y., & Ibadurrahman, I. (2022). Prediksi Kualitas Kopi Dengan Algoritma Random Forest Melalui Pendekatan Data Science. *Data Sciences Indonesia (DSI)*, 2(1), 20–29.
- Dachi, A. (2020). Inovasi Produk terhadap Keputusan Pembelian dan dampaknya terhadap loyalitas pelanggan: Studi Pengguna Mobil Toyota Calya di Kota Bogor dan Bekasi. *JSHP: Jurnal Sosial Humaniora Dan Pendidikan*, 4(2), 120–129.
- Devianto, Y., & Dwiasnati, S. (2020). Kerangka kerja sistem kecerdasan buatan dalam meningkatkan kompetensi sumber daya manusia Indonesia. *IncomTech: Jurnal Telekomunikasi Dan Komputer*, 10(1), 19–24.
- Dewi, A. F., & Pratiwi, R. (2021). Analisis regresi logistik biner pada pengaruh harga, kualitas pelayanan dan promosi terhadap kepuasan pelanggan dalam menggunakan jasa layanan grab di kabupaten lamongan. *Inferensi*, 4(2), 77–84.

- Dewi, P., Aulia, R. N., & Taufiqillah, R. (2024). Customer Churn Prediction for Life Insurance Using Binary Logistic Regression. *Economic Reviews Journal*, 3(3), 2289–2299.
- Dwiastanti, A., & others. (2022). Pengaruh kualitas produk dan harga terhadap loyalitas pelanggan pada amstirdam coffee di malang. *INSPIRASI: JURNAL ILMU-ILMU SOSIAL*, 19(1), 600–610.
- Fahreza, M. R., & Putra, A. A. (2024). Crm di era media sosial: Memanfaatkan platform sosial untuk meningkatkan interaksi pelanggan. *Musytari: Neraca Manajemen, Akuntansi, Dan Ekonomi*, 6(1), 61–70.
- Gill-Wiehl, A., Miles, S., Wu, J., & Kammen, D. M. (2022). Beyond customer acquisition: A comprehensive review of community participation in mini grid projects. *Renewable and Sustainable Energy Reviews*, 153, 111778.
- Handayani, F. (2021). Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 7(3), 329–334.
- Hidayat, R., Syawaludin, M. A., & Nurmalitasari, N. (2024). Prediksi Churn Pelanggan Multinational Bank Menggunakan Algoritma Machine Learning. *Simpatik: Jurnal Sistem Informasi Dan Informatika*, 4(2), 89–97.
- Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. *Briefings in Bioinformatics*, 24(2), bbad002.
- Iranzad, R., & Liu, X. (2024). A review of random forest-based feature selection methods for data science education and applications. *International Journal of Data Science and Analytics*, 1–15.
- Junaedy, N. A. A., & Asfiah, N. (2024). Hubungan antara AI, Machine Learning,

- Dan Implikasinya Terhadap Responsivitas Bisnis. *Jurnal Bisnis Inovatif Dan Digital*, 1(3), 81–91.
- Lestari, A., Santoso, R., & Suparti, S. (2024). ANALISIS SENTIMEN PENGGUNA ONLINE TRAVEL AGENT (OTA) PADA PERUSAHAAN PEGIPEGI. COM MENGGUNAKAN RANDOM FOREST. *Jurnal Gaussian*, 12(4), 616–624.
- Mahira, M., Hadi, P., & Nastiti, H. (2021). Pengaruh kualitas produk dan kualitas pelayanan terhadap kepuasan pelanggan Indihome. *Konferensi Riset Nasional Ekonomi Manajemen Dan Akuntansi*, 2(1), 1267–1283.
- Manongga, D., Rahardja, U., Sembiring, I., Lutfiani, N., & Yadila, A. B. (2022). RETRACTED (Di Tarik): Dampak Kecerdasan Buatan Bagi Pendidikan. *ADI Bisnis Digital Interdisiplin Jurnal*, 3(2), 110–124.
- Manzali, Y., & Elfar, M. (2023). Random forest pruning techniques: a recent review. *Operations Research Forum*, 4(2), 43.
- Martiwi, R. (2021). Implementasi Relationship Marketing dalam Industri Pariwisata dan Perhotelan. *Komitmen: Jurnal Ilmiah Manajemen*, 2(1), 21–28.
- Rahmat, W. A., Ladjamuddin, S. M., & Awaludin, D. T. (2023). Perbandingan Algoritma Decision Tree, Random Forest dan Naive Bayes pada Prediksi Penilaian Kepuasan Penumpang Maskapai Pesawat Menggunakan Dataset Kaggle. *Jurnal Rekayasa Informasi*, 12(2), 150–159.
- Rizky, A. T. (2023). *PENERAPAN NAÏVE BAYES CLASSIFIER UNTUK ANALISIS SENTIMEN DATA ULASAN APLIKASI E-COMMERCE SHOPEE PADA GOOGLE PLAY STORE*. Universitas Siliwangi.
- Siregar, M. F. A., & others. (2024). PENGARUH PEMASARAN MEDIA SOSIAL

TERHADAP KESADARAN MEREK DAN AKUISISI KONSUMEN
MEREK MAKANAN CEPAT SAJI PADA GENERASI Y DAN Z DI
JAKARTA. *Jurnal Ekonomi Trisakti*, 4(1), 57–68.

Sotarjua, L. M., & Santoso, D. B. (2022). Perbandingan Algoritma Knn, Decision Tree,* Dan Random* Forest Pada Data Imbalanced Class Untuk Klasifikasi Promosi Karyawan. *Jurnal INSTEK (Informatika Sains Dan Teknologi)*, 7(2), 192–200.

Usman, S., Aziz, F., & others. (2023). Analisis Perilaku Pelanggan menggunakan Metode Ensemble Logistic Regression. *JURNAL TEKNOLOGI DAN ILMU KOMPUTER PRIMA (JUTIKOMP)*, 6(2), 90–97.

Wahyudi, W., & Sutoyo, S. (2022). Penerapan customer acquisition dalam pertumbuhan bisnis pada Dapoer Super Sambal Padang. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 8(2), 109–115.

Wahyudia, I. W. T., & Kadyanana, I. G. A. G. A. (2023). Implementasi Logistic Regression Dalam Sistem Diagnosa Penyakit Diabetes Dengan KNN. *Jurnal Elektronik Ilmu Komputer Udayana P-ISSN*, 2301, 5373.

Watugilang, A., & Heikal, J. (2024). Pengaruh Kualitas Jasa Servis terhadap Kepuasan Pelanggan Perusahaan Servis Kalibrasi Alat Survey Geomatika di Jakarta dengan Binary Logistic Regression. *Indonesian Research Journal on Education*, 4(4), 79–83.

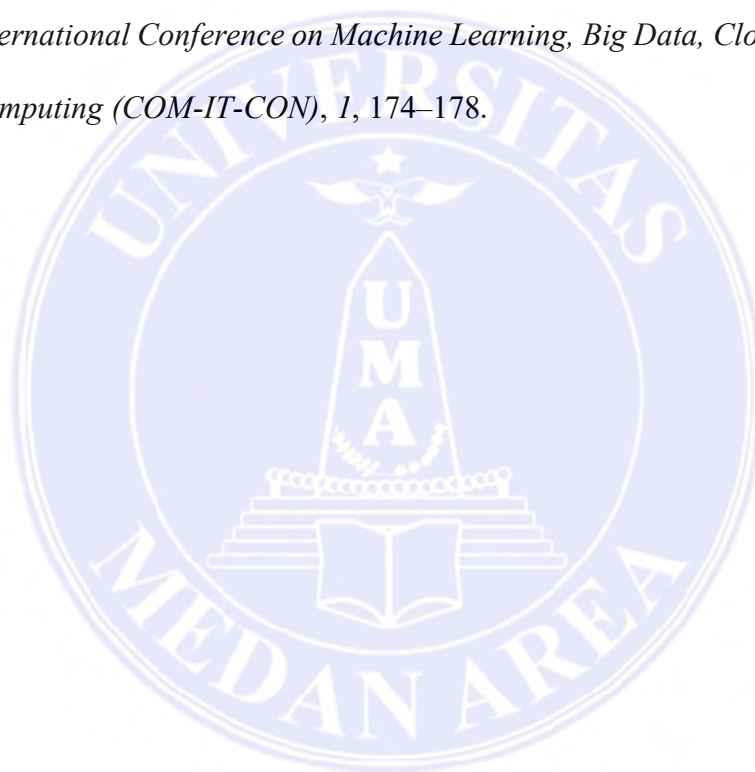
Widjanarko, W., Saputra, F., & Hadita, H. (2023). Analisis Keputusan Pembelian dan Loyalitas Pelanggan terhadap Voucher Gratis Ongkos Kirim E-Commerce Shopee Indonesia. *Jurnal Ilmu Manajemen Terapan*, 4(5), 678–685.

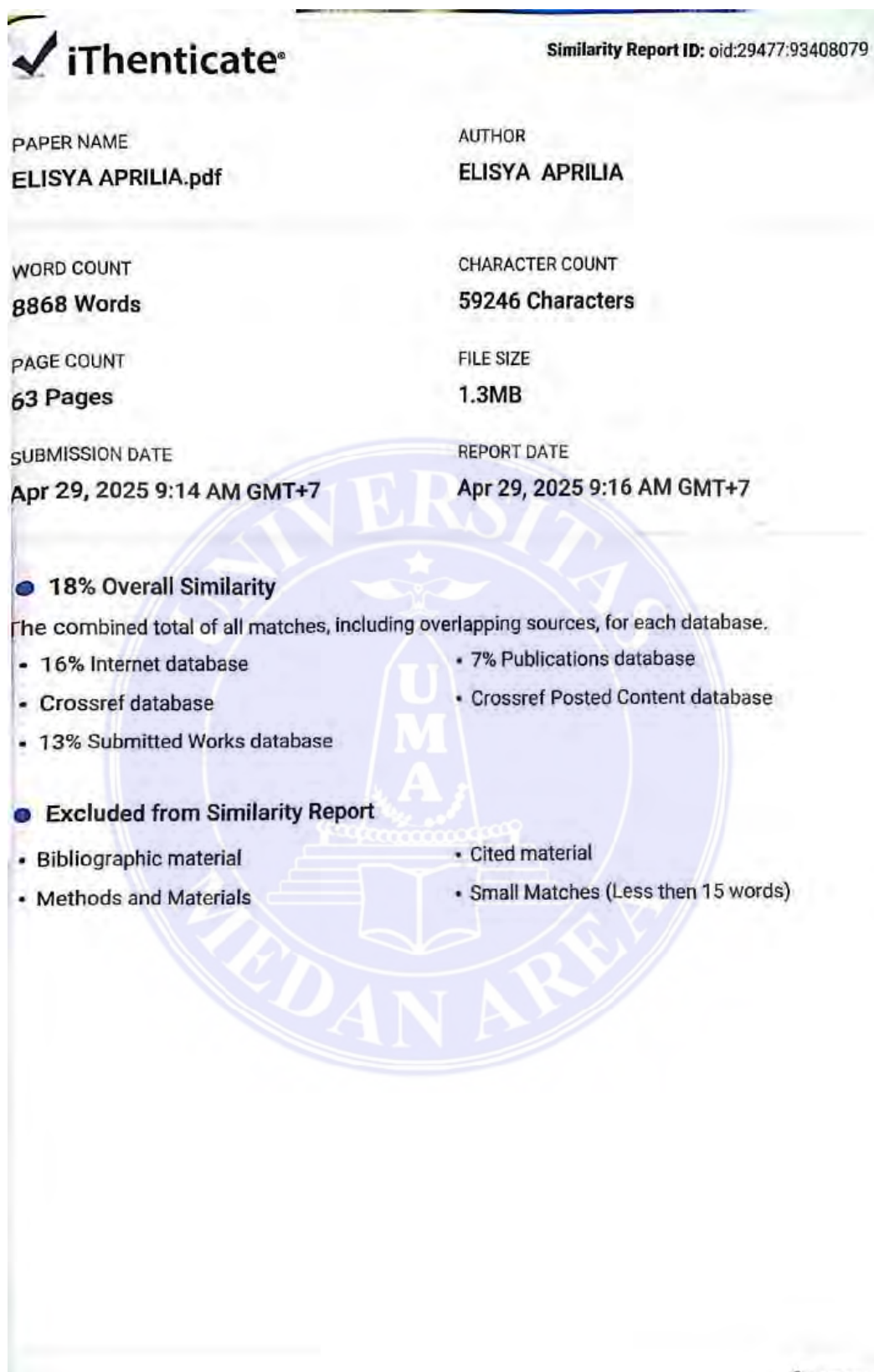
Wilson, J. R., Lorenz, K. A., & Selby, L. P. (2024). Introduction to binary logistic

regression. *Modeling Binary Correlated Responses: Using SAS, SPSS, R and STATA*, 3–18.

Yang, Z., Diao, Z., & Kang, J. (2020). Customer management in Internet-based platform firms: review and future research directions. *Marketing Intelligence & Planning*, 38(7), 957–973.

Zadoo, A., Jagtap, T., Khule, N., Kedari, A., & Khedkar, S. (2022). A review on churn prediction and customer segmentation using machine learning. *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, 1, 174–178.





iThenticate® Similarity Report ID: oid:29477:93408079

PAPER NAME	AUTHOR
ELISYA APRILIA.pdf	ELISYA APRILIA

WORD COUNT	CHARACTER COUNT
8868 Words	59246 Characters

PAGE COUNT	FILE SIZE
63 Pages	1.3MB

SUBMISSION DATE	REPORT DATE
Apr 29, 2025 9:14 AM GMT+7	Apr 29, 2025 9:16 AM GMT+7

18% Overall Similarity

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

- 16% Internet database
- 7% Publications database
- Crossref database
- Crossref Posted Content database
- 13% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Cited material
- Methods and Materials
- Small Matches (Less than 15 words)



UNIVERSITAS MEDAN AREA
FAKULTAS TEKNIK
Kampus I : Jalan Kolam Nomor 1 Medan Estate / Jalan Gedung PBSI, Medan 20223
Kampus II : Jalan Sei Seraya Nomor 70 A / Jalan Sei A Buall Nomor 79 B, Medan 20112 Telepon : (061) 8225602, 8201994
Fax : (061) 8226331 IIP : 0811 607 259 website: www.uma.ac.id Email : univ_medanarea@uma.ac.id

Nomor : 2719/FT/01.10/IX/2024

25 September 2024

Lampiran : -

Hal : Pembimbing Tugas Akhir

Yth. Pembimbing Tugas Akhir
ANDRE HASUDUNGAN LBS S.Ti, M.Sc (Sebagai Pembimbing)
di Tempat

Dengan hormat, sehubungan telah dipenuhinya persyaratan untuk memperoleh Tugas Akhir dari mahasiswa atas :

Nama : ELISYA APRILIA
NIM : 218160016
Jurusan : TEKNIK INFORMATIKA

Maka dengan hormat kami mengharapkan kesediaan saudara :

ANDRE HASUDUNGAN LBS S.Ti, M.Sc (Sebagai Pembimbing)

Adapun Tugas Akhir Skripsi berjudul :

**PERBANDINGAN PERFORMA DARI LOGISTIC REGRESSION DAN RANDOM FOREST
DALAM MEMPREDIKSI AKUISISI PELANGGAN**

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.

Dekan,



Dr Eng. Supriatno.ST, MT.





UNIVERSITAS MEDAN AREA

FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate ☎ (061) 7360168, Medan, 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 42402994, Medan, 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 456 /FT.6/01.10/XI/2024 25 November 2024
Lamp : -
Hal : Penelitian Dan Pengambilan Data Tugas Akhir

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

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

NO	N A M A	N P M	PRODI
1	Elisya Aprilia	218160016	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 semata-mata untuk tujuan Ilmiah dan Skripsi, yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul :

Perbandingan Performa dan Logistic Regression dan Random Forest dalam Memprediksi Akuisisi Pelanggan.

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.


Dr. Firdausy Priatno, ST, MT

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

