

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

**PENERAPAN DATA MINING UNTUK KLASIFIKASI
LOYALITAS DAN PENGGAN E-COMMERCE
SHOPEE MENGGUNAKAN ALGORITMA
GRAFIKAL BOOSTING**



SKRIPSI

**Diajukan Sebagai Syarat Untuk Menyelesaikan Pendidikan
Program Sarjana (S-1)
Pada Program Studi Sistem Informasi**

**Oleh :
FERI BASOFI
NIM : 2102030046**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU TEKNIK
UNIVERSITAS BINA INSAN
2025**

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

HALAMAN PENGESAHAN SKRIPSI



PENERAPAN DATA MINING UNTUK KLASIFIKASI

LOYALITAS PELANGGAN *E-COMMERCE*

SHOPEE MENGGUNAKAN ALGORITMA

GRADIENT BOOSTING

Oleh:

Feri Basofi

NIM : 2102030046

Lubuklinggau, Januari 2025

Pembimbing I

Pembimbing II

Budi Santoso, M.Kom

Bunga Intan, M.Kom

**Mengetahui,
Dekan Fakultas Ilmu Teknik
Universitas Bina Insan Lubuklinggau**

Dr. Rudi Kurniawan, ST M.Kom

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)
HALAMAN PERSETUJUAN PENGUJI



Pada hari Tanggal ... Bulan telah dilaksanakan Ujian Skripsi oleh Program Studi sistem informasi Fakultas Ilmu Teknik Universitas Bina Insan.

Nama : Feri Basofi

Nim : 2102030046

Judul Skripsi : Penerapan Data mining untuk Klasifikasi Loyalitas Pelanggan
E-Commerce Shopee menggunakan Algoritma *Gradient Boosting*

Komisi penguji

1. Ketua : Budi Santoso, M.Kom (.....)
2. Sekretaris : Bunga Intan, M.Kom (.....)
3. Anggota : Joni Karman, M.Kom (.....)

Mengetahui,

**Kepala Program Studi Sistem Informasi
Fakultas Ilmu Teknik
Universitas Bina insan**

(Harma Oktafia Lingga Wijaya, M,Kom)

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

HALAMAN MOTO DAN PERSEMBAHAN



MOTTO :

- ✦ *Ketekunan dan usaha keras adalah kunci kesuksesan.*
- ✦ *Setiap langkah mendekatkan kita pada impian.*

Persembahan Kepada :

- ❖ *Ibu dan Bapak tercinta, yang telah banyak mendukungku dan memberikan do'a untuk keberhasilanku*
- ❖ *Diriku Sendiri*
- ❖ *Rekan kerja, yang selalu mendukung dan menjadi sumber inspirasi*
- ❖ *Teman-teman seperjuanganku*
- ❖ *Almaterku.*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

HALAMAN PERNYATAAN



Saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Feri Basofi
NIM : 2102030046
Program Studi : Sistem Informasi
Fakultas : Ilmu Teknik

Menyatakan dengan sesungguhnya bahwa penelitian dan penulisan Skripsi yang saya susun sebagai persyaratan untuk memperoleh gelar Sarjana (S-1) Universitas Bina Insan, merupakan hasil kerja saya sendiri dan tidak menyuruh orang lain yang mengerjakannya. Ada pun bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain dan telah saya tuliskan sumbernya secara jelas sesuai dengan norma, kaidah dan etika penulisan ilmiah.

Jika dikemudian hari ternyata terbukti bahwa penelitian dan tugas akhir ini bukan hasil kerja saya sendiri atau plagiat dalam bagian-bagian tertentu, maka saya bersedia dikenakan sanksi sesuai dengan peraturan perundangan yang berlaku.

Lubuklinggau Januari 2025

Penulis,

Feri Basofi

NIM.2102030046

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

ABSTRACT




Indonesia's e-commerce industry is growing rapidly, with Shopee as the market leader. This research analyzes the classification of customer loyalty on the Shopee e-commerce platform with the Gradient Boosting Decision Tree (GBDT) algorithm. The goal is to identify loyalty levels based on data collected through the Shopee API, focusing on transactions and user behavior. The data obtained will help develop a classification model to distinguish loyal customers and not, followed by a pre-processing stage to clean the data. The research method used in this study is a quantitative approach with an exploratory and descriptive design, which aims to understand, explore, and classify the loyalty of Shopee customers based on their customer characteristics and transaction data. The quantitative approach was chosen because it provides the ability to analyze the relationship between various variables that can be measured numerically. Based on the model evaluation table which presents metrics such as Precision, Recall, F1-Score, and Accuracy, the results obtained for all categories (Loyal and Non-Loyal) reached 100%. This indicates that the applied classification model has excellent performance in recognizing these categories without error. All predictions generated by the model, both positive and negative, match the actual values. This study successfully identified the key factors that influence Shopee customer loyalty in Musi Rawas. Using the Gradient Boosting algorithm, the developed classification model shows 100% accuracy in categorizing customers. The dominant factors found were Average Total Payment, Customer Group, and Purchase Amount. These results provide insight into customer behavior and can be used to improve Shopee's marketing strategy.

Keywords: customer loyalty, gradient boosting, shopee e-commerce, data classification

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

ABSTRAK



Industri *e-commerce* Indonesia tumbuh pesat, dengan Shopee sebagai pemimpin pasar. Penelitian ini menganalisis klasifikasi loyalitas pelanggan pada *e-commerce* Shopee dengan algoritma *Gradient Boosting*. Tujuannya adalah mengidentifikasi tingkat loyalitas berdasarkan data yang dikumpulkan melalui API Shopee, fokus pada transaksi dan perilaku pengguna. Data yang diperoleh akan membantu mengembangkan model klasifikasi untuk membedakan pelanggan loyal dan tidak, diikuti dengan tahap pra-pemrosesan untuk membersihkan data. Metode penelitian yang digunakan dalam penelitian ini adalah pendekatan kuantitatif dengan desain eksploratori dan deskriptif, yang bertujuan untuk memahami, menggali, dan mengklasifikasikan loyalitas pelanggan Shopee berdasarkan karakteristik pelanggan dan data transaksi mereka. Pendekatan kuantitatif dipilih karena memberikan kemampuan untuk menganalisis hubungan antara berbagai variabel yang dapat diukur secara numerik. Berdasarkan tabel evaluasi model yang menyajikan metrik seperti *Precision*, *Recall*, *F1-Score*, dan *Accuracy*, hasil yang diperoleh untuk semua kategori (Loyal dan Tidak Loyal) mencapai 100%. Hal ini menandakan bahwa model klasifikasi yang diterapkan memiliki kinerja yang sangat baik dalam mengenali kategori tersebut tanpa kesalahan. Semua prediksi yang dihasilkan oleh model, baik yang positif maupun negatif, sesuai dengan nilai yang sebenarnya. Penelitian ini berhasil mengidentifikasi faktor-faktor kunci yang mempengaruhi loyalitas pelanggan Shopee di Musi Rawas. Dengan menggunakan algoritma *Gradient Boosting*, model klasifikasi yang dikembangkan menunjukkan akurasi 100% dalam mengkategorikan pelanggan. Faktor dominan yang ditemukan adalah Total Pembayaran Rata-rata, Kelompok Pelanggan, dan Jumlah Pembelian. Hasil ini memberikan wawasan tentang perilaku pelanggan dan dapat digunakan untuk meningkatkan strategi pemasaran Shopee.

Kata kunci: loyalitas pelanggan, *gradient boosting*, *e-commerce* shopee, klasifikasi data

Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)
KATA PENGANTAR



Alhamdulillah puji dan penulis ucapkan kepada Allah SWT atas segala rahmat dan karunia-Nya memberikan kekuatan dan kesempatan, sehingga penulis dapat menyelesaikan skripsi ini dengan maksimal, Untuk diajukan sebagai syarat menyelesaikan pendidikan program Sarjana (S-1) Pada Program Studi Sistem Informasi Fakultas Ilmu Teknik Universitas Bina Insan. Sholawat beserta salam semoga tetap tercurahkan kepada bagi Nabi Muhammad SAW, keluarga, sahabat, serta umatnya hingga akhir zaman. Selama proses penulisan dan penyusunan skripsi ini, penulis telah berusaha sebaik-baiknya untuk dapat menyelesaikan skripsi ini baik tepat pada waktunya. Penulis menyadari bahwa skripsi ini tentunya masih jauh dari sempurna dan mungkin terdapat kesalahan baik sengaja maupun tidak sengaja. Oleh karena itu, kritik dan saran yang membangun tentunya sangat diharapkan dari berbagai pihak. Penulis mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu selama proses penyelesaian skripsi ini diantaranya yaitu:

1. Bapak Dr. H. Sardiyo, M.M. selaku Rektor Universitas Bina Insan.
2. Bapak Dr. Muhamad Akbar, S.T., M.IT selaku Wakil Rektor I Universitas Bina Insan.
3. Bapak Wakhid Nur Mukhlis, M.Pd., M.M selaku Wakil Rektor II Universitas Bina Insan.
4. Bapak Dr. Rudi Kurniawan, ST M.Kom selaku Dekan Fakultas Ilmu Teknik Universitas Bina Insan yang telah banyak memberikan bimbingan dan arah dalam penulisan Skripsi ini.
5. Ibu Harma Oktafia Lingga Wijaya, M,Kom selaku Kepala Program studi Sistem Informasi Fakultas Ilmu Teknik Universitas Bina Insan yang telah banyak memberikan bimbingan dan arah dalam penulisan Skripsi ini.
6. Bapak Budi Santoso, M.Kom selaku Pembimbing I yang telah banyak memberikan bimbingan dan arah dalam penulisan Skripsi ini.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

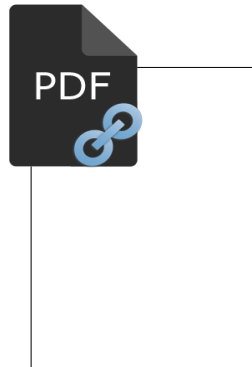
7. Ibu Bunga Intan, M.Kom selaku Pembimbing II yang telah banyak memberikan bimbingan dan arahan dalam penulisan Skripsi ini.
8. Bapak Joni Karman, M.Pd selaku Penguji yang telah banyak memberikan bimbingan dan arahan dalam penulisan Skripsi ini.
9. Seluruh Staf Dosen dan Karyawan Universitas Bina Insan Lubuklinggau yang telah banyak memberikan ilmu pengetahuan dan bimbingan kepada penulis.
10. Teristimewa untuk kedua orangtua saya, Ayahanda Dasiman dan Ibunda Samini yang telah memberikan kasih sayangnya, perhatian, dan dukungan moril maupun materil tanpa henti bagi penulis serta menguatkan penulis dalam doa-doanya. Orangtua motivator terbesar saya untuk terus melangkah meraih mimpi-mimpi akan masa depan dan Orangtua yang sangat luar biasa.
11. Seluruh teman-teman Prodi Sistem Informasi angkatan 2021 yang tidak dapat disebutkan satu persatu .

Akhir kata semoga penelitian ini dapat bermanfaat bagi untuk penelitian selanjutnya.

Lubuklinggau,.....Januari 2025

Penulis

Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)
DAFTAR RIWAYAT HIDUP



Biodata

Nama : Feri Basofi
Tempat / Tanggal Lahir : Kalibening, 13 Mei 1998
Jenis Kelamin : Laki-laki
Agama : Islam
Alamat : Desa Kalibening Kec. Tugumulyo Kab.
Musi Rawas

Pendidikan

- SD : SD Negeri Kalibening
- SMP/MTS Sederajat : SMP Negeri L.Sidoharjo
- SMA/MAN/SMK Sederajat : SMK Negeri Tugumulyo

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR ISI



Halaman

HALAMAN PENGESAHAN SK	i
HALAMAN PERSETUJUAN PENGUJI	ii
HALAMAN MOTO DAN PERSEMBAHAN	iii
HALAMAN PERNYATAAN	iv
ABSTRACT	v
ABSTRAK	vi
KATA PENGANTAR	vii
DAFTAR RIWAYAT HIDUP	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Penelitian	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	4
1.4 Batasan Masalah	4
1.5 Tujuan dan Manfaat Penelitian	5
1.6 Sistematika Penulisan	8
BAB II TINJAUAN PUSTAKA	9
2.1 Literatur	9
2.2 Penelitian Terdahulu Yang Relevan	27
2.3 Kerangka Berpikir	32
BAB III METODELOGI PENELITIAN	34
3.1 Metode Penelitian	34
3.2 Metode Pengumpulan Data	34
3.3 Metode Analisa	35
3.4 Tempat dan Waktu Penelitian	37
3.5 Alat dan Bahan	38
3.6 Metode Pengujian dan Pengolahan Data	38
BAB IV	44
HASIL DAN PEMBAHASAN	44
4.1 Gambaran Umum	44
4.2 Hasil	45
4.2.1 Pengumpulan Data	45
4.2.2 Preprocessing	48
4.2.3 Labelling	52
4.2.4 Visualisasi	53
4.2.5 Klasifikasi	56
4.2.6 Evaluasi	58
4.3 Pembahasan	68

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

BAB V KESIMPULAN DAN SARAN	70
5.1. Kesimpulan	70
5.2. Saran	71
DAFTAR PUSTAKA	72
LAMPIRAN	74



Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR TABEL


	Halaman
Tabel 2.1 Penelitian Terdahulu.....	27
Tabel 3.1 Jadwal Penelitian.....	37
Tabel 3.2 Alat.....	38
Tabel 3.3 Bahan	38
Tabel 4.1 Label Loyalitas.....	68
Tabel 4.2 Klasifikasi Gradient Boosting.....	68



Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR GAMBAR

	Halaman
 Gambar 2.1 Kerangka Berpikir	32
Gambar 3.1 <i>Flowchart</i> Pengujian Model	39
Gambar 3.2 <i>Flowchart</i> Pengolahan Data	41
Gambar 4.1 <i>Pseudocode</i> Pengambilan Data.....	48
Gambar 4.2 Hasil Pengambilan Data Pelanggan.....	48
Gambar 4.3 Hasil Pengambilan Data Transaksi.....	48
Gambar 4.4 Hasil Pembersihan Data.....	49
Gambar 4.5 <i>Pseudocode</i> Pembersihan Data.....	49
Gambar 4.6 Hasil Normalisasi Data	50
Gambar 4.7 <i>Pseudocode</i> Normalisasi Data	50
Gambar 4.8 Hasil <i>Feature Engineering</i>	51
Gambar 4.9 <i>Pseudocode Feature Engineering</i>	52
Gambar 4.10 Hasil <i>Labelling</i>	52
Gambar 4.11 <i>Pseudocode Labelling</i>	53
Gambar 4.12 Hasil Komposisi Loyalitas.....	53
Gambar 4.13 Diagram Loyalitas	54
Gambar 4.14 Grafik <i>Boxplot</i>	54
Gambar 4.15 Grafik <i>Parplot</i>	55
Gambar 4.16 <i>Pseudocode gradient boosting</i>	58
Gambar 4.17 Hasil Evaluasi <i>Classification Report</i>	59
Gambar 4.18 Hasil <i>Confusion Matrix</i>	59
Gambar 4.19 Diagram <i>Feature Importance</i>	59
Gambar 4.20 Diagram <i>Heatmap</i>	60
Gambar 4.21 Grafik <i>Partial Dependence Plot</i>	61
Gambar 4.22 Diagram Loyalitas Pelanggan.....	62
Gambar 4.23 Diagram Loyalitas Berdasarkan Kategori Produk	63
Gambar 4.24 Diagram <i>Scatter Plot</i>	64
Gambar 4.25 <i>Boxplot</i> perbandingan	65
Gambar 4.26 Diagram Distribusi Kelompok Usia	66
Gambar 4.27 Diagram Loyalitas Berdasarkan Kelompok Pelanggan.....	67

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR LAMPIRAN



	Halaman
Lampiran 1 Pengesahan Judul.....	75
Lampiran 2 Lembar Bimbingan.....	76
Lampiran 3 Dataset.....	78

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

BAB I



1.1 Latar Belakang Penelitian

Perkembangan industri *e-commerce* di Indonesia telah mengalami lonjakan yang signifikan dalam beberapa tahun terakhir. Sebagai salah satu negara dengan jumlah pengguna internet terbesar di Asia Tenggara, Indonesia memiliki potensi pasar *e-commerce* yang sangat besar. Data dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menunjukkan bahwa per Januari 2023[1], jumlah pengguna internet di Indonesia mencapai lebih dari 200 juta orang, yang sebagian besar di antaranya terhubung dengan platform *e-commerce* seperti Shopee, Tokopedia, Bukalapak, dan Lazada. Di antara platform-platform ini, Shopee mencatatkan pertumbuhan yang sangat pesat dan menjadi salah satu pilihan utama konsumen Indonesia dalam melakukan transaksi *online*. Berdasarkan laporan iPrice pada kuartal ketiga 2022, Shopee berada di posisi pertama dalam hal jumlah kunjungan bulanan di Indonesia, dengan lebih dari 200 juta kunjungan per bulan, mengungguli para pesaingnya. Keberhasilan ini tidak terlepas dari banyaknya inovasi yang dilakukan Shopee, baik dari sisi fitur aplikasi, promosi, maupun pengembangan sistem logistik yang memudahkan konsumen dalam berbelanja.

Namun, meskipun Shopee terus berkembang pesat, tantangan yang dihadapi oleh platform ini adalah bagaimana menjaga loyalitas pelanggan agar tetap setia menggunakan layanan mereka, meskipun banyak pilihan *e-commerce* lainnya yang tersedia[2]. Loyalitas pelanggan menjadi faktor krusial dalam mempertahankan pangsa pasar dan meningkatkan pendapatan jangka panjang. Tingkat loyalitas pelanggan pada platform *e-commerce* di Indonesia, khususnya Shopee, telah menjadi salah satu indikator penting dalam mengevaluasi keberhasilan strategi bisnis digital. Berdasarkan laporan umum dari Statista 2023, Shopee secara konsisten memimpin pasar *e-commerce* di Indonesia dalam hal jumlah pengguna aktif dan total pesanan per kuartal. Namun, tantangan tetap ada,

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

terutama dalam meningkatkan retensi pelanggan di tengah persaingan ketat *e-commerce* regional dan lokal (Gusti, 2023). Untuk wilayah Sumatera Selatan, belum ada data spesifik yang tersedia secara luas, sehingga diperlukan penelitian lebih lanjut untuk menggambarkan tingkat loyalitas pelanggan di daerah seperti Musi Rawas.

Banyak faktor yang mempengaruhi keputusan seseorang untuk terus berbelanja di platform tertentu, termasuk pengalaman pengguna, kualitas layanan pelanggan, harga produk, kecepatan pengiriman, serta program loyalitas atau *reward* yang ditawarkan oleh platform *e-commerce*[3]. Di sisi lain, dengan semakin berkembangnya data yang tersedia, *e-commerce* kini memiliki kesempatan untuk memanfaatkan teknik data mining dalam menggali pola perilaku pelanggan yang sulit terlihat secara langsung. Salah satu teknik data mining yang banyak digunakan untuk klasifikasi dan prediksi adalah algoritma *Gradient Boosting*. Algoritma ini termasuk dalam kategori *ensemble learning* yang menggabungkan beberapa model prediksi sederhana untuk membentuk model yang lebih kuat.

Salah satu wilayah yang memiliki dinamika pasar yang menarik untuk dianalisis dalam konteks ini adalah Musi Rawas, yang terletak di Provinsi Sumatera Selatan. Sebagai daerah yang masih dalam tahap perkembangan dari sisi digitalisasi, Musi Rawas memberikan tantangan tersendiri dalam hal penetrasi dan adopsi *e-commerce*. Meskipun kota-kota besar seperti Jakarta, Bandung, dan Surabaya mendominasi transaksi *e-commerce* di Indonesia, Musi Rawas, yang lebih dikenal dengan potensi sektor pertanian dan sumber daya alamnya, menunjukkan tren yang menarik dalam adopsi teknologi dan penggunaan *e-commerce*. Berdasarkan data yang dirilis oleh Badan Pusat Statistik (BPS) Provinsi Sumatera Selatan pada 2022, penetrasi internet di daerah ini meningkat pesat, terutama setelah pandemi COVID-19 yang mempercepat digitalisasi di berbagai sektor. Hal ini tercermin dalam semakin banyaknya masyarakat yang mulai berbelanja secara *online*, khususnya melalui platform *e-commerce* seperti Shopee. Namun, masih banyak faktor yang mempengaruhi loyalitas pelanggan *e-*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

e-commerce, seperti faktor geografis dan sosial yang mempengaruhi pola perilaku belanja, seperti tingkat pendapatan, aksesibilitas internet, serta kebiasaan berbelanja tradisional, dapat mempengaruhi keputusan pelanggan dalam menggunakan layanan *e-commerce*.

Dalam konteks *e-commerce* Shopee, pelanggan di Musi Rawas memiliki beragam karakteristik yang mempengaruhi loyalitas mereka. Salah satu faktor yang signifikan adalah kebiasaan berbelanja *online* yang masih berkembang. Meskipun penetrasi internet semakin meningkat, beberapa pelanggan di daerah ini mungkin masih memilih berbelanja secara langsung di pasar tradisional atau toko fisik, terutama untuk kebutuhan sehari-hari.[5] Faktor lain yang mempengaruhi loyalitas adalah daya beli, yang sering kali lebih rendah dibandingkan dengan pelanggan di kota-kota besar. Hal ini membuat harga produk dan program diskon atau cashback menjadi sangat penting dalam mempertahankan pelanggan. Dengan demikian, analisis yang mendalam menggunakan teknik data mining dapat membantu Shopee untuk lebih memahami perilaku pelanggan di Musi Rawas, serta merancang strategi yang lebih sesuai dengan kondisi pasar lokal.

Penelitian ini tidak hanya berfokus pada penerapan teknik data mining untuk mengklasifikasikan loyalitas pelanggan Shopee menggunakan algoritma *Gradient Boosting*, tetapi juga mencoba memberikan gambaran yang lebih luas tentang bagaimana platform *e-commerce* seperti Shopee dapat memanfaatkan data yang ada untuk meningkatkan pengalaman pelanggan dan memperkuat loyalitas mereka, terutama di daerah seperti Musi Rawas yang memiliki karakteristik pasar yang unik. Dengan penerapan metode ini, Shopee diharapkan dapat lebih memahami perilaku konsumen dan merancang strategi yang lebih efektif untuk menjaga loyalitas pelanggan di era persaingan *e-commerce* yang semakin ketat.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

1.2 Identifikasi Masalah

- a. Salah satu masalah utama penelitian ini adalah bagaimana menentukan faktor-faktor yang paling signifikan untuk mengklasifikasikan loyalitas pelanggan di Shopee. Apakah frekuensi pembelian, nilai transaksi, jenis produk yang dibeli, atau interaksi pelanggan dengan fitur aplikasi yang lebih mempengaruhi loyalitas.
- b. Pelanggan di Musi Rawas mungkin memiliki perilaku belanja yang berbeda karena faktor akses internet dan daya beli, yang perlu dipertimbangkan dalam model klasifikasi loyalitas.
- c. Implementasi model klasifikasi dalam dunia nyata harus dapat berkontribusi pada strategi pemasaran yang lebih efektif dan meningkatkan retensi pelanggan.

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah yang telah dijelaskan, rumusan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

- a. Bagaimana cara mengidentifikasi faktor-faktor utama yang mempengaruhi loyalitas pelanggan Shopee di Musi Rawas menggunakan data transaksi dan perilaku pelanggan?
- b. Bagaimana penerapan algoritma *Gradient Boosting* dalam mengklasifikasikan tingkat loyalitas pelanggan Shopee berdasarkan data transaksi dan perilaku mereka?

1.4 Batasan Masalah

Agar penelitian ini dapat lebih terfokus dan terarah, berikut adalah batasan-batasan yang akan diterapkan dalam penelitian ini:

- a. Penelitian ini akan difokuskan pada pelanggan Shopee yang berada di wilayah Musi Rawas, Sumatera Selatan. Penelitian ini tidak mencakup pelanggan Shopee dari daerah lain atau seluruh Indonesia.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- b. Data yang digunakan dalam penelitian ini hanya akan mencakup data transaksi pelanggan Shopee yang tersedia dalam sistem Shopee dan relevan dengan penelitian ini. Data yang dianalisis mencakup informasi tentang riwayat pembelian, frekuensi transaksi, jenis produk yang dibeli, serta interaksi pelanggan dengan berbagai fitur yang ada di platform Shopee.
- c. Penelitian ini akan membatasi penggunaan algoritma *machine learning* pada algoritma **Gradient Boosting** sebagai metode utama untuk klasifikasi loyalitas pelanggan. loyalitas pelanggan dalam penelitian ini akan diklasifikasikan dalam dua kategori: **Loyal** (pelanggan yang sering bertransaksi), **Tidak Loyal** (pelanggan yang jarang bertransaksi atau berhenti bertransaksi dalam periode tertentu).

1.5 Tujuan dan Manfaat Penelitian

- a. Tujuan Penelitian

Penelitian ini bertujuan untuk:

- 1) Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor utama yang mempengaruhi loyalitas pelanggan Shopee, khususnya pelanggan yang berada di wilayah Musi Rawas. Faktor-faktor ini meliputi frekuensi pembelian, nilai transaksi, jenis produk yang dibeli, serta interaksi pelanggan dengan fitur-fitur dalam aplikasi Shopee.
- 2) Penelitian ini bertujuan untuk mengembangkan model klasifikasi loyalitas pelanggan Shopee berdasarkan data transaksi dan perilaku pelanggan dengan menggunakan algoritma *Gradient Boosting*. Model ini diharapkan dapat mengklasifikasikan pelanggan Shopee ke dalam kategori loyal, tidak loyal, berdasarkan pola perilaku yang teridentifikasi dalam data.
- 3) Penelitian ini bertujuan untuk mengevaluasi sejauh mana model klasifikasi yang dibangun dapat membantu Shopee dalam merancang strategi pemasaran yang lebih personal dan meningkatkan loyalitas pelanggan.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Penelitian ini juga akan mengukur dampak praktis dari penerapan model dalam meningkatkan retensi pelanggan dan konversi penjualan.

b. Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat bagi berbagai pihak, baik secara teoritis maupun praktis:

1) Manfaat Teoritis

- a) Penelitian ini akan memberikan kontribusi pada pengembangan ilmu pengetahuan di bidang data mining, khususnya dalam penerapan algoritma *Gradient Boosting* untuk klasifikasi loyalitas pelanggan *e-commerce*. Temuan dari penelitian ini dapat digunakan sebagai referensi bagi penelitian lebih lanjut yang membahas penggunaan algoritma ini dalam sektor *e-commerce* di Indonesia.
- b) Penelitian ini juga akan memperkaya literatur mengenai model klasifikasi loyalitas pelanggan dalam konteks *e-commerce*. Model yang dikembangkan dapat dijadikan dasar untuk penelitian serupa yang bertujuan untuk memahami loyalitas pelanggan di platform *e-commerce* lainnya.

2) Manfaat Praktis

- a) Hasil penelitian ini diharapkan dapat memberikan wawasan bagi Shopee dalam merancang strategi pemasaran yang lebih efektif dan personal, dengan menggunakan data transaksi dan perilaku pelanggan. Dengan memahami faktor-faktor yang mempengaruhi loyalitas pelanggan, Shopee dapat mengoptimalkan penawaran produk, promosi, dan program loyalitas untuk meningkatkan retensi pelanggan.
- b) Penelitian ini akan memberikan rekomendasi yang berguna bagi Shopee untuk meningkatkan pemahaman mereka terhadap pasar di Musi Rawas, serta perbedaan perilaku yang ada dibandingkan dengan kota-kota besar. Hal ini penting untuk merancang strategi yang lebih sesuai

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

dengan karakteristik lokal dan meningkatkan penetrasi pasar di wilayah tersebut.

- c) Hasil penelitian ini dapat mendorong perusahaan *e-commerce* lainnya untuk memanfaatkan algoritma data mining, khususnya *Gradient Boosting*, untuk memahami perilaku pelanggan dan meningkatkan loyalitas mereka. Dengan demikian, penelitian ini dapat membantu memperkuat penerapan teknologi analitik di sektor *e-commerce* Indonesia.

3) Manfaat Sosial

- a) Penelitian ini berpotensi memberikan manfaat sosial, terutama bagi pelanggan Shopee yang berada di daerah-daerah terpencil seperti Musi Rawas. Dengan adanya pemahaman yang lebih mendalam tentang perilaku belanja mereka, Shopee dapat mengoptimalkan layanan dan penawaran agar lebih sesuai dengan kebutuhan dan preferensi pelanggan di wilayah tersebut. Ini akan meningkatkan pengalaman berbelanja yang lebih baik dan dapat mempercepat adopsi *e-commerce* di daerah tersebut.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

1.6 Sistematika Penulisan

Dalam penulisan Skripsi yang merupakan laporan dari hasil penelitian, direncanakan terdiri dari lima bab. Masing-masing bab berisi :

BAB I : PENDAHULUAN

Dalam bab ini berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, dan sistematika penulisan.

BAB II : KAJIAN PUSTAKA

Dalam bab ini berisi teori-teori yang mendasari masalah yang diteliti.

BAB III : METODOLOGI PENELITIAN

Dalam bab ini berisi tentang metode penelitian, metode pengumpulan data, metode analisa, tempat dan waktu penelitian, alat dan bahan, metode pengujian dan pengolahan data.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil dan pembahasan.

BAB V : KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari seluruh penelitian dan saran-saran/masukan-masukan yang berguna dimasa yang akan datang.

DAFTAR PUSTAKA

LAMPIRAN



2.1 Literatur

2.1.1 Klasifikasi

Klasifikasi adalah salah satu teknik utama dalam data mining yang digunakan untuk mengelompokkan data ke dalam kategori-kategori tertentu berdasarkan atribut atau fitur yang relevan.[6] Proses ini penting dalam banyak aplikasi, terutama dalam pengambilan keputusan berbasis data yang semakin berkembang, seperti di bidang *e-commerce*. Dalam konteks *e-commerce*, klasifikasi dapat digunakan untuk mengidentifikasi pola atau tren yang muncul dari perilaku pelanggan, dan salah satu contohnya adalah pengklasifikasian loyalitas pelanggan. Loyalitas pelanggan adalah faktor kunci yang mempengaruhi keberlanjutan dan pertumbuhan suatu bisnis, terutama dalam dunia *e-commerce* yang sangat kompetitif seperti Shopee.

Shopee, sebagai salah satu platform *e-commerce* terbesar di Asia Tenggara, memiliki jutaan pelanggan yang aktif berbelanja setiap hari. Dalam rangka untuk mempertahankan pelanggan setia dan meningkatkan pengalaman mereka, Shopee perlu menganalisis perilaku pelanggan secara mendalam. Oleh karena itu, klasifikasi loyalitas pelanggan menjadi sangat penting. Dengan mengklasifikasikan pelanggan ke dalam kategori loyal atau tidak loyal, Shopee dapat merancang strategi yang lebih efektif untuk meningkatkan loyalitas dan retensi pelanggan.

Selain itu, penggunaan klasifikasi dalam konteks loyalitas pelanggan *e-commerce* juga membawa dampak positif dalam hal pengelolaan sumber daya. Shopee dapat memfokuskan lebih banyak perhatian dan sumber daya untuk mempertahankan pelanggan yang loyal, misalnya dengan memberikan penawaran khusus, diskon, atau program loyalitas yang lebih menarik. Di sisi lain, pelanggan

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

yang dianggap tidak loyal atau yang berisiko berhenti berbelanja dapat diberi perhatian lebih sedikit atau disentif untuk kembali berinteraksi dengan platform. Dengan demikian, platform dapat meningkatkan loyalitas pelanggan ini membantu Shopee dalam pengambilan keputusan yang lebih efisien dan berbasis data.

Selain itu, hasil klasifikasi ini dapat digunakan untuk segmentasi pasar yang lebih tepat, yang penting untuk strategi pertumbuhan Shopee di pasar yang sangat kompetitif. *E-commerce* sering kali beroperasi dengan volume data yang sangat besar, dan menggunakan algoritma seperti *Gradient Boosting* memungkinkan untuk memanfaatkan data tersebut secara maksimal untuk menciptakan wawasan yang lebih mendalam mengenai preferensi pelanggan. Wawasan ini memungkinkan Shopee untuk lebih memahami perilaku pelanggan mereka, termasuk faktor-faktor yang mempengaruhi loyalitas dan keputusan pembelian.

Klasifikasi menggunakan *Gradient Boosting* dalam konteks loyalitas pelanggan Shopee menawarkan banyak manfaat, tidak hanya untuk mengidentifikasi pelanggan yang loyal dan tidak loyal, tetapi juga untuk merancang strategi yang lebih efektif dalam mempertahankan pelanggan. Dengan kemampuannya dalam menangani data yang kompleks dan menghasilkan prediksi yang akurat, *Gradient Boosting* memberikan alat yang sangat berguna dalam memahami perilaku pelanggan, yang pada akhirnya membantu Shopee untuk mengoptimalkan pengalaman pelanggan dan meningkatkan keberhasilan bisnis mereka.

2.1.2 Loyalitas Pelanggan

Loyalitas pelanggan merupakan salah satu aspek yang paling fundamental dalam dunia bisnis, terutama dalam sektor *e-commerce* yang sangat dinamis.[7] Loyalitas pelanggan merujuk pada kecenderungan pelanggan untuk tetap berinteraksi dengan merek atau platform dalam jangka panjang, yang tercermin melalui pembelian berulang, kepuasan, serta rekomendasi positif yang diberikan kepada orang lain. Dalam dunia *e-commerce*, loyalitas pelanggan sangat vital,

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

karena pelanggan yang loyal tidak hanya berkontribusi pada pendapatan berkelanjutan, tetapi juga merekomendasikan merek yang memperluas pangsa pasar melalui *word of mouth* atau dari mulut ke mulut. Oleh karena itu, pemahaman mendalam tentang loyalitas pelanggan dan faktor-faktor yang mempengaruhinya menjadi sangat penting bagi perusahaan *e-commerce* seperti Shopee untuk tetap bersaing di pasar yang semakin kompetitif.

Loyalitas pelanggan di *e-commerce* dapat diukur melalui berbagai indikator yang menunjukkan seberapa besar komitmen pelanggan terhadap platform tersebut. Salah satu indikator utama adalah frekuensi pembelian, di mana pelanggan yang loyal cenderung melakukan pembelian secara berulang kali dalam periode waktu tertentu. Sebaliknya, pelanggan yang tidak loyal mungkin hanya melakukan satu kali pembelian dan kemudian berhenti berinteraksi dengan platform tersebut. Frekuensi pembelian ini bisa dijadikan acuan untuk menilai seberapa sering seorang pelanggan menggunakan platform *e-commerce* untuk memenuhi kebutuhan mereka. Selain itu, nilai transaksi juga merupakan indikator penting dalam menilai loyalitas pelanggan. Pelanggan yang sering melakukan pembelian dengan nilai transaksi yang signifikan menunjukkan bahwa mereka tidak hanya puas dengan produk atau layanan yang ditawarkan, tetapi juga merasa cukup percaya diri untuk melakukan pembelian dalam jumlah besar. Dengan kata lain, semakin besar nilai transaksi pelanggan, semakin besar kemungkinan mereka untuk menjadi pelanggan yang loyal.

Namun, loyalitas pelanggan tidak hanya terbatas pada angka transaksi dan frekuensi pembelian. Interaksi dengan fitur-fitur yang ada di platform *e-commerce* juga dapat menjadi indikator penting dalam mengukur loyalitas. Fitur-fitur seperti program loyalitas, rekomendasi produk, atau diskon khusus sering kali memengaruhi keputusan pelanggan untuk terus berinteraksi dengan platform tersebut. Pelanggan yang sering memanfaatkan fitur-fitur ini menunjukkan bahwa mereka tidak hanya menggunakan platform untuk berbelanja, tetapi juga merasa terikat dengan berbagai fitur tambahan yang meningkatkan pengalaman mereka. Misalnya, pelanggan yang sering menggunakan fitur-fitur seperti kupon diskon atau mengikuti program reward pelanggan biasanya memiliki tingkat loyalitas

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

yang lebih tinggi dibandingkan pelanggan yang jarang menggunakan fitur tersebut.

Loyalitas pelanggan juga dikategorikan menjadi dua jenis, yaitu loyalitas afektif dan loyalitas konatif. Loyalitas afektif merujuk pada hubungan emosional yang terjalin antara pelanggan dan merek atau platform, di mana pelanggan merasa puas dan terikat secara emosional dengan produk atau layanan yang diberikan. Hal ini dapat terjadi karena pengalaman pelanggan yang positif, layanan yang baik, atau bahkan faktor eksternal lainnya seperti nilai sosial yang dimiliki oleh perusahaan. Pelanggan yang loyal secara afektif seringkali akan mengabaikan harga yang lebih tinggi atau beralih ke pesaing meskipun ada tawaran yang lebih murah, karena mereka merasa puas dan terhubung secara emosional dengan merek tersebut. Di sisi lain, loyalitas konatif merujuk pada niat atau kecenderungan pelanggan untuk melakukan pembelian kembali di masa depan. Loyalitas konatif ini lebih berfokus pada perilaku yang dapat diamati secara langsung, seperti pengulangan pembelian atau perencanaan untuk membeli produk dari merek yang sama.

Loyalitas pelanggan juga sangat terkait dengan kepuasan pelanggan. Pelanggan yang puas dengan pengalaman mereka berbelanja di suatu platform *e-commerce* cenderung lebih setia. Kepuasan ini bisa berasal dari berbagai faktor, termasuk kualitas produk, kemudahan navigasi platform, layanan pelanggan yang responsif, atau kecepatan pengiriman. Shopee, misalnya, memiliki berbagai fitur untuk meningkatkan pengalaman pelanggan, seperti pengiriman gratis, sistem pembayaran yang aman, dan layanan pelanggan yang responsif, yang semuanya berkontribusi pada tingkat kepuasan pelanggan. Kepuasan ini menjadi dasar bagi terciptanya loyalitas pelanggan, karena pelanggan yang puas lebih cenderung untuk kembali berbelanja dan bahkan merekomendasikan platform tersebut kepada orang lain.

Salah satu cara yang efektif untuk meningkatkan loyalitas pelanggan di *e-commerce* adalah dengan menerapkan program loyalitas. Program ini biasanya menawarkan berbagai insentif kepada pelanggan yang melakukan pembelian secara teratur atau yang menghabiskan jumlah tertentu dalam pembelian. Program

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

loyalitas seperti diskon, poin *reward*, atau bahkan penghargaan berupa hadiah dapat membuat pelanggan merasa dihargai dan memotivasi mereka untuk tetap menggunakan platform tersebut. Platform juga mengimplementasikan program-program seperti *Shopee Coins* yang memungkinkan pelanggan mengumpulkan poin dari setiap transaksi yang dapat ditukarkan dengan diskon atau produk gratis.[8] Program seperti ini tidak hanya meningkatkan kepuasan pelanggan, tetapi juga mendorong mereka untuk berbelanja lebih sering.


Loyalitas pelanggan juga bisa dipengaruhi oleh pengalaman pelanggan secara keseluruhan saat berinteraksi dengan platform. Dalam *e-commerce*, pengalaman pelanggan mencakup segala sesuatu mulai dari navigasi situs web, kemudahan pencarian produk, kualitas deskripsi produk, hingga kecepatan layanan pengiriman. Pengalaman yang buruk dalam satu atau lebih aspek ini dapat merusak loyalitas pelanggan dan menyebabkan mereka beralih ke platform *e-commerce* lain. Sebaliknya, pengalaman berbelanja yang menyenangkan dan tanpa hambatan akan memperkuat hubungan antara pelanggan dan platform tersebut. Oleh karena itu, pengelolaan pengalaman pelanggan yang optimal adalah kunci dalam menjaga loyalitas pelanggan.

Pada akhirnya, loyalitas pelanggan adalah fondasi yang sangat penting bagi kelangsungan hidup dan pertumbuhan bisnis *e-commerce*. Pelanggan yang loyal tidak hanya akan terus membeli produk atau layanan, tetapi juga akan membawa keuntungan jangka panjang melalui promosi gratis yang mereka lakukan dengan cara merekomendasikan platform kepada orang lain. Oleh karena itu, perusahaan *e-commerce* seperti *Shopee* harus terus berupaya untuk menjaga dan meningkatkan loyalitas pelanggan mereka, dengan menawarkan produk dan layanan terbaik, memperhatikan pengalaman pelanggan, serta memanfaatkan teknologi untuk menciptakan hubungan yang lebih erat dan personal dengan pelanggan. Dalam pasar yang sangat kompetitif saat ini, loyalitas pelanggan adalah faktor yang tidak dapat diabaikan dan menjadi kunci utama untuk bertahan dan tumbuh di industri *e-commerce*.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

2.1.3 *E-commerce*



E-commerce, atau perdagangan elektronik, adalah sebuah konsep yang merujuk pada transaksi jual beli produk atau layanan yang dilakukan secara elektronik melalui internet. Dalam beberapa dekade terakhir, *e-commerce* telah berkembang pesat dan menjadi bagian integral dari kehidupan sehari-hari, baik bagi konsumen maupun bagi pelaku bisnis.[9] Dengan memanfaatkan teknologi internet, *e-commerce* memungkinkan perusahaan dan individu untuk melakukan transaksi tanpa batasan geografis, waktu, dan ruang. Perkembangan ini memberikan dampak yang sangat besar, tidak hanya bagi model bisnis tradisional, tetapi juga bagi cara orang berbelanja, dan melakukan interaksi dengan produk atau layanan.

Salah satu keuntungan utama dari *e-commerce* adalah kemampuannya untuk menjangkau audiens yang lebih luas, baik di tingkat lokal maupun global. Sebelumnya, bisnis yang mengandalkan transaksi fisik memerlukan keberadaan fisik di lokasi tertentu, yang sering kali membatasi kemampuan mereka untuk menjangkau konsumen yang lebih luas. Dengan adanya *e-commerce*, perusahaan dapat menjual produk dan layanan mereka kepada konsumen di seluruh dunia, selama mereka memiliki koneksi internet. Ini membuka peluang baru untuk pasar global, yang sebelumnya mungkin sulit dijangkau oleh bisnis berskala kecil hingga menengah. Dengan cara ini, *e-commerce* memungkinkan perusahaan untuk mengembangkan jangkauan pasar mereka secara signifikan dan meningkatkan potensi pendapatan.

Di Indonesia, *e-commerce* telah berkembang sangat pesat dengan hadirnya berbagai platform besar seperti Shopee, Tokopedia, dan Bukalapak, yang memfasilitasi pembelian dan penjualan barang melalui platform *online*. Platform-platform ini menawarkan berbagai macam produk, mulai dari barang-barang kebutuhan sehari-hari hingga barang-barang mewah, dari kategori *fashion* hingga elektronik, dan dari makanan hingga peralatan rumah tangga. Dengan menyediakan akses mudah ke berbagai produk dan layanan, platform-platform *e-commerce* ini telah menggantikan banyak fungsi toko fisik, menjadikannya

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

pilihan utama bagi konsumen yang ingin berbelanja dengan lebih praktis, cepat, dan efisien. Shopee, misalnya, telah berhasil menarik perhatian konsumen Indonesia dengan berbagai fitur promo menarik, pengiriman gratis, serta pilihan pembayaran yang beragam.

Namun, *e-commerce* bukan hanya soal transaksi jual beli. Ia juga menciptakan ekosistem yang mengubah cara orang berinteraksi dengan produk dan layanan. Salah satu faktor penting yang membedakan *e-commerce* dengan model bisnis tradisional adalah kemudahan akses yang ditawarkan kepada konsumen. Konsumen kini dapat membeli produk atau layanan kapan saja dan di mana saja, hanya dengan beberapa klik atau ketukan di layar ponsel mereka. Berbelanja menjadi lebih fleksibel dan tidak terbatas pada jam operasional toko fisik. Fitur-fitur seperti pencarian produk, rekomendasi personal, dan ulasan dari pembeli lain juga memudahkan konsumen dalam mengambil keputusan pembelian. Semua kemudahan ini mengarah pada peningkatan kenyamanan dan kepuasan konsumen, yang pada gilirannya berdampak pada loyalitas mereka terhadap platform *e-commerce* tertentu.

Salah satu perkembangan terbaru dalam dunia *e-commerce* adalah penggunaan teknologi canggih untuk meningkatkan pengalaman berbelanja. Misalnya, banyak platform *e-commerce* kini memanfaatkan *big data* dan kecerdasan buatan (*AI*) untuk menganalisis perilaku konsumen dan memberikan rekomendasi produk yang lebih personal. Dengan menggunakan data transaksi dan perilaku pelanggan, platform *e-commerce* dapat mengidentifikasi preferensi dan kebiasaan belanja konsumen, kemudian menawarkan produk yang sesuai dengan kebutuhan mereka. Hal ini membuat pengalaman belanja menjadi lebih relevan dan menarik, serta meningkatkan kemungkinan pembelian berulang.

Selain itu, perkembangan teknologi pembayaran digital juga memainkan peran besar dalam kemajuan *e-commerce*. Dengan adanya berbagai pilihan pembayaran digital seperti *e-wallet*, kartu kredit, hingga transfer bank, proses pembayaran dalam *e-commerce* menjadi lebih mudah, aman, dan cepat. Integrasi teknologi pembayaran ini tidak hanya mempermudah transaksi bagi konsumen,

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

tetapi juga memberikan kenyamanan dan fleksibilitas dalam cara mereka membayar, yang semakin mendorong platform *e-commerce* dengan konsumen.

Salah satu tantangan terbesar platform *e-commerce* adalah persaingan yang ketat. Dengan semakin banyaknya platform *e-commerce* yang bermunculan, setiap perusahaan dituntut untuk selalu menawarkan sesuatu yang unik dan berbeda agar dapat memenangkan hati konsumen. Selain kualitas produk dan pelayanan, platform *e-commerce* juga perlu memperhatikan faktor seperti kecepatan pengiriman, kemudahan penggunaan aplikasi atau website, dan berbagai fitur menarik yang dapat membuat konsumen merasa nyaman dan puas. Kompetisi ini mendorong inovasi yang berkelanjutan dalam industri *e-commerce*, yang pada gilirannya memberikan lebih banyak pilihan dan manfaat bagi konsumen.

E-commerce telah merevolusi cara masyarakat melakukan aktivitas berbelanja dan berbisnis, dan terus berkembang seiring dengan kemajuan teknologi.[10] Platform seperti Shopee, Tokopedia, dan Bukalapak menjadi contoh nyata dari bagaimana *e-commerce* dapat memberikan kemudahan, kenyamanan, dan pilihan yang luas bagi konsumen. Dengan kualitas pelayanan yang baik, pengalaman pengguna yang menyenangkan, dan penggunaan teknologi canggih, platform *e-commerce* dapat membangun loyalitas pelanggan yang kuat dan mempertahankan posisi mereka dalam pasar yang semakin kompetitif. Ke depannya, *e-commerce* akan terus berkembang dan menjadi bagian tak terpisahkan dari ekosistem ekonomi global, dengan semakin banyak inovasi yang akan mempermudah hidup konsumen dan meningkatkan kualitas layanan.

2.1 Shopee

Shopee telah menjadi salah satu platform *e-commerce* terbesar dan paling populer di Asia Tenggara, termasuk Indonesia. Sejak diluncurkan pada tahun 2015, Shopee telah berhasil menciptakan ekosistem belanja *online* yang inovatif dan menghubungkan penjual dengan pembeli secara langsung melalui platform yang mudah digunakan. Shopee menawarkan berbagai kategori produk yang

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

mencakup hampir semua kebutuhan konsumen, mulai dari pakaian, peralatan rumah tangga, elektronik, hingga makanan dan kecantikan. Keberagaman produk yang ditawarkan, ditambah dengan harga yang kompetitif, telah menjadikan Shopee sebagai salah satu pilihan utama bagi konsumen di Indonesia dan di negara-negara Asia Tenggara lainnya. Keberhasilan Shopee tidak hanya terletak pada ketersediaan produk yang lengkap dan harga yang terjangkau, tetapi juga pada strategi pemasaran yang efektif dan pengalaman pelanggan yang memuaskan.

2.2 Data mining

Data mining merupakan proses yang sangat penting dalam dunia analisis data modern, yang bertujuan untuk mengekstraksi informasi yang berguna dari dataset besar dan kompleks. Seiring dengan perkembangan teknologi informasi, volume data yang tersedia bagi organisasi dan perusahaan semakin meningkat, menjadikan data mining sebagai salah satu alat utama untuk menganalisis dan memahami data secara lebih mendalam.[11] Proses data mining melibatkan penggunaan teknik-teknik analisis yang canggih untuk menemukan pola, tren, atau hubungan tersembunyi dalam data yang mungkin tidak terlihat secara langsung. Dalam banyak kasus, *data mining* dapat membantu mengubah data mentah menjadi wawasan yang berharga, yang dapat digunakan untuk mendukung pengambilan keputusan bisnis yang lebih cerdas dan berbasis bukti.

Secara umum, *data mining* melibatkan berbagai tahap yang mulai dari pengumpulan data, pembersihan data, hingga analisis dan interpretasi hasil. Salah satu hal yang membuat data mining begitu penting adalah kemampuannya untuk menangani dataset yang sangat besar dan beragam, seperti yang ditemukan dalam industri *e-commerce*. Platform *e-commerce* seperti Shopee mengumpulkan sejumlah besar data pelanggan setiap harinya, termasuk data transaksi, perilaku pengguna, preferensi produk, dan banyak lagi. Data ini bisa sangat beragam dan tidak terstruktur, sehingga sulit untuk dipahami atau dianalisis tanpa teknik yang tepat. Di sinilah *data mining* berperan penting: dengan menggunakan algoritma

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

dan teknik analisis yang tepat, *data mining* memungkinkan penemuan wawasan yang dapat meningkatkan pemahaman tentang pelanggan serta memperbaiki pengalaman mereka.

Keuntungan utama dan penggunaan *data mining* dalam konteks *e-commerce* adalah kemampuannya untuk mengungkap pola-pola tersembunyi yang mungkin tidak terlihat dengan analisis tradisional. Di dalam platform *e-commerce* seperti Shopee, data yang terkumpul bisa sangat kompleks dan beragam, dengan banyak faktor yang memengaruhi keputusan pembelian pelanggan. Tanpa *data mining*, sulit untuk menganalisis data dalam skala besar dan menemukan hubungan atau tren yang mungkin ada di balik perilaku pelanggan. Misalnya, *data mining* dapat membantu Shopee mengidentifikasi apakah pelanggan yang lebih sering membeli produk di kategori tertentu lebih cenderung menggunakan voucher atau program cashback, atau apakah ada pola musiman dalam pembelian produk yang dapat dimanfaatkan dalam perencanaan stok dan promosi.

Data mining merupakan alat yang sangat kuat dalam analisis data untuk bisnis *e-commerce*, khususnya dalam memahami dan mengelompokkan pelanggan berdasarkan perilaku dan loyalitas mereka. Dengan memanfaatkan teknik *data mining* seperti klasifikasi, *regresi*, dan *clustering*, Shopee dapat merancang strategi yang lebih tepat dan personal untuk meningkatkan pengalaman pelanggan, menjaga loyalitas mereka, dan meningkatkan pendapatan perusahaan. Dalam dunia yang semakin didorong oleh data, kemampuan untuk mengekstrak wawasan dari data pelanggan adalah kunci untuk bertahan dan berkembang dalam industri *e-commerce* yang sangat kompetitif.

2.3 Gradient Boosting

Gradient Boosting adalah salah satu algoritma pembelajaran mesin yang sangat populer dan efektif dalam memecahkan berbagai masalah klasifikasi dan regresi.[12] Algoritma ini termasuk dalam kategori *ensemble learning*, yaitu pendekatan yang menggabungkan beberapa model prediksi sederhana untuk menciptakan model yang lebih kuat dan akurat. Pada dasarnya, *ensemble learning*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

berfokus pada ide bahwa kombinasi dari beberapa model prediksi yang lemah dapat menghasilkan model yang jauh lebih baik daripada model tunggal. Salah satu alasan mengapa *Gradient Boosting* itu efektif adalah kemampuannya dalam mengurangi kesalahan prediksi secara iteratif dan memberikan bobot lebih pada kesalahan yang lebih besar, sehingga model dapat fokus pada data yang lebih sulit diprediksi. Dengan cara ini, algoritma ini dapat memperbaiki prediksi secara progresif dan lebih mendalam pada setiap iterasi.

Salah satu fitur utama dari *Gradient Boosting* adalah bagaimana algoritma ini mengatasi kesalahan atau residual error pada setiap iterasi. Pada setiap langkah, model berusaha untuk memperbaiki kesalahan prediksi yang terjadi pada data training.[13] Ini dilakukan dengan menghitung gradien dari kesalahan prediksi dan kemudian mengoptimalkan parameter model untuk mengurangi kesalahan tersebut. Rumus berikut menggambarkan langkah penting dalam menghitung bobot pembaruan pada setiap iterasi:

$$\text{Bobot Pembaruan} = \frac{\sum \text{Residual}_i}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)]}$$

Rumus ini menunjukkan bahwa setiap pembaruan dihitung berdasarkan residual (kesalahan prediksi sebelumnya) dan probabilitas sebelumnya, dengan mempertimbangkan distribusi data. Proses ini membantu model fokus pada prediksi yang lebih sulit, sehingga meningkatkan akurasi secara iteratif. Dengan cara ini, *Gradient Boosting* dapat belajar lebih mendalam dari data, memperbaiki kesalahan prediksi, dan menghasilkan model yang sangat kuat dan akurat.

2.4 Dataset

Dataset merupakan elemen fundamental dalam setiap proyek *machine learning*. Tanpa *dataset* yang relevan dan berkualitas, model yang dibangun tidak akan mampu memberikan hasil yang akurat dan bermanfaat.[14] Secara sederhana, *dataset* adalah kumpulan data yang digunakan untuk melatih dan

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

menguji model *machine learning*. Proses pelatihan model melibatkan penggunaan *dataset* untuk mengajarkan model mengenali pola dan hubungan dalam data, sedangkan pengujian model dilakukan untuk menilai seberapa baik model tersebut bekerja dengan data yang berbeda-beda yang pernah dilihat sebelumnya. Kualitas dan representativitas *dataset* sangat berpengaruh pada kinerja akhir model, yang menjadikannya salah satu tahap yang paling kritis dalam setiap proyek *machine learning*. Dalam konteks penelitian ini, dataset yang digunakan berfokus pada data transaksi pelanggan di platform *e-commerce* Shopee.

2.5 API Shopee

API (Application Programming Interface) Shopee merupakan salah satu komponen penting dalam dunia *e-commerce* yang memungkinkan akses langsung ke berbagai data yang ada di platform Shopee, termasuk data transaksi, informasi produk, dan data pengguna. Sebagai salah satu platform *e-commerce* terbesar di Asia Tenggara, Shopee menyediakan *API* yang memungkinkan para pengembang untuk berinteraksi dengan sistem mereka dan mengambil data secara otomatis tanpa perlu mengakses antarmuka pengguna secara langsung. Dengan menggunakan *API* ini, data yang diperlukan dapat diambil dengan cara yang lebih efisien dan lebih terstruktur. Dalam konteks penelitian ini, *API* Shopee digunakan untuk mengumpulkan data pelanggan dan transaksi yang akan dianalisis untuk tujuan klasifikasi loyalitas pelanggan.[15]

Salah satu fitur menarik dari *API* Shopee adalah kemampuannya untuk memberikan data tentang interaksi pengguna dengan aplikasi Shopee, yang mencakup aktivitas pelanggan dalam menggunakan berbagai fitur, seperti voucher diskon, promo cashback, dan program loyalitas lainnya. Data ini sangat berharga karena interaksi pelanggan dengan fitur-fitur ini sering kali menjadi indikator penting dalam menentukan loyalitas pelanggan. Sebagai contoh, pelanggan yang sering menggunakan voucher atau yang aktif dalam mengikuti promo diskon besar-besaran mungkin lebih cenderung untuk terus melakukan pembelian berulang, yang menunjukkan tanda-tanda loyalitas. Dengan *API* Shopee, data

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

terkait penggunaan voucher dan promo dapat dikumpulkan secara otomatis, memungkinkan peneliti untuk menganalisis pengaruh dari program-program ini terhadap loyalitas pelanggan.



2.6 Labelling

Labelling adalah proses fundamental dalam setiap tugas klasifikasi yang melibatkan *machine learning*, di mana data diberikan label atau penandaan yang sesuai berdasarkan karakteristik tertentu. Dalam konteks penelitian ini, *labelling* dilakukan untuk mengklasifikasikan pelanggan Shopee ke dalam dua kategori utama: loyal dan tidak loyal. Proses penandaan data ini merupakan langkah awal yang sangat penting karena tanpa adanya label yang tepat, model *machine learning* tidak akan mampu mempelajari pola yang relevan dalam data dan menghasilkan prediksi yang akurat. Dalam penelitian yang bertujuan untuk mengklasifikasikan loyalitas pelanggan, *labelling* menjadi dasar dari segala proses analisis lebih lanjut, karena menentukan bagaimana data akan diproses dan dipelajari oleh algoritma klasifikasi.

Labelling juga memainkan peran yang sangat penting dalam evaluasi model yang dibangun. Setelah model dilatih menggunakan data yang sudah dilabeli, hasil prediksi model akan dibandingkan dengan label asli untuk mengukur kinerja model tersebut. Salah satu cara yang umum digunakan untuk mengevaluasi kinerja model klasifikasi adalah dengan menggunakan *confusion matrix*, yang menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas. *Confusion matrix* ini membantu dalam memahami sejauh mana model berhasil dalam mengklasifikasikan pelanggan menjadi loyal atau tidak loyal, dan di mana model mungkin mengalami kesalahan. Oleh karena itu, ketepatan *labelling* akan mempengaruhi kualitas evaluasi model dan dapat memberikan petunjuk bagi pengembang untuk melakukan perbaikan atau penyesuaian pada algoritma.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

2.7 Preprocessing Data

Preprocessing data adalah langkah krusial dalam setiap proyek data *science* atau *machine learning*. Fungsi untuk menyiapkan data agar siap dianalisis atau digunakan dalam model.[16] Tanpa langkah *preprocessing* yang tepat, data yang digunakan bisa mengandung kesalahan, ketidaksesuaian, atau informasi yang hilang yang dapat mengganggu hasil analisis atau model yang dibangun. Dalam konteks penelitian ini, *preprocessing* data memainkan peranan penting untuk memastikan bahwa data yang digunakan dalam algoritma *Gradient Boosting* bersih, terstruktur dengan baik, dan siap untuk analisis yang lebih mendalam. Proses ini mencakup berbagai tahap, termasuk pembersihan data, transformasi data, normalisasi, serta penanganan *missing values*, yang kesemuanya sangat vital untuk meningkatkan kualitas dan akurasi model *machine learning* yang akan digunakan dalam klasifikasi loyalitas pelanggan Shopee.

2.8 Confusion Matrix

Confusion Matrix adalah alat yang sangat penting dalam dunia *machine learning*, khususnya dalam evaluasi kinerja model klasifikasi. Tabel ini menyediakan gambaran yang jelas mengenai bagaimana model melakukan prediksi terhadap data yang sebenarnya, dengan mengkategorikan hasil prediksi menjadi beberapa kategori yang berbeda. Dalam konteks penelitian ini, yang berfokus pada klasifikasi loyalitas pelanggan Shopee, *confusion matrix* sangat berguna untuk menilai kemampuan model dalam memprediksi loyalitas pelanggan, baik yang loyal maupun tidak loyal, berdasarkan data yang ada. Dengan menggunakan *confusion matrix*, Kemampuan algoritma klasifikasi dapat diukur untuk menilai tingkat akurasi, seperti *Gradient Boosting*, bekerja dalam membuat prediksi yang tepat untuk pelanggan, dan seberapa banyak kesalahan yang terjadi selama proses klasifikasi.

Secara umum, *confusion matrix* terdiri dari empat elemen utama yang menggambarkan hasil klasifikasi, yaitu *True Positive (TP)*, *True Negative (TN)*,

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

False Positive (FP), dan *False Negative* (FN). *True Positive* (TP) merujuk pada jumlah kasus di mana model benar mengklasifikasikan pelanggan yang sebenarnya loyal sebagai loyal. Sementara itu, *True Negative* (TN) adalah jumlah kasus di mana model dengan benar mengklasifikasikan pelanggan yang sebenarnya tidak loyal sebagai tidak loyal. Sementara itu, *False Positive* (FP) menunjukkan jumlah kasus di mana model salah mengklasifikasikan pelanggan yang sebenarnya tidak loyal sebagai loyal, dan *False Negative* (FN) adalah jumlah kasus di mana model salah mengklasifikasikan pelanggan yang sebenarnya loyal sebagai tidak loyal. Keempat elemen ini memberikan informasi yang sangat berharga mengenai jenis kesalahan yang dilakukan oleh model dan area mana yang perlu diperbaiki atau ditingkatkan.

Salah satu metrik yang sangat bergantung pada *confusion matrix* adalah *accuracy*, yang mengukur seberapa sering model membuat prediksi yang benar. *Accuracy* dihitung dengan rumus:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Metrik ini memberikan gambaran umum tentang kinerja model, tetapi hanya mengandalkan jumlah prediksi yang benar tanpa mempertimbangkan jenis kesalahan yang terjadi. Oleh karena itu, meskipun *accuracy* memberikan informasi yang berguna, tidak selalu cukup untuk menilai kinerja model secara menyeluruh, terutama dalam situasi di mana data yang tidak seimbang (misalnya, lebih banyak pelanggan yang tidak loyal dibandingkan pelanggan loyal) dapat mempengaruhi hasilnya.

Untuk mengatasi kekurangan *accuracy*, Metrik lain yang lebih spesifik juga dapat digunakan untuk evaluasi, seperti *precision*, *recall*, dan *F1-score*. *Precision* mengukur sejauh mana model menghindari kesalahan positif, yaitu seberapa banyak dari prediksi pelanggan yang loyal yang benar-benar loyal. *Precision* dihitung dengan rumus:

$$Precision = \frac{TP}{TP + FP}$$

Precision sangat berguna ketika biaya kesalahan prediksi positif (misalnya, mengklasifikasikan pelanggan yang tidak loyal sebagai loyal) cukup

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

besar. Dalam hal ini, penting untuk memastikan bahwa ketika model mengidentifikasi seseorang sebagai loyal, Kemungkinan besar individu tersebut benar-benar loyal. *Precision* yang tinggi berarti bahwa model sangat hati-hati dalam memberikan label "loyal" hanya melakukannya ketika sangat yakin.

Di sisi lain, *recall* mengukur seberapa banyak pelanggan yang benar-benar loyal dapat ditemukan oleh model, yaitu sejauh mana model berhasil mengidentifikasi pelanggan loyal yang sebenarnya. *Recall* dihitung dengan rumus:

$$Recall = \frac{TP}{TP + FN}$$

Recall menjadi metrik yang penting ketika tujuan utama adalah memastikan model tidak melewatkan pelanggan yang loyal, meskipun ini mungkin menyebabkan kesalahan dalam mengklasifikasikan beberapa pelanggan yang tidak loyal sebagai loyal. Dalam penelitian ini, *recall* berfungsi untuk memaksimalkan tingkat keberhasilan dalam mengidentifikasi pelanggan loyal, yang penting untuk program-program retensi pelanggan atau strategi pemasaran yang ditujukan untuk meningkatkan loyalitas.

Namun, *precision* dan *recall* sering kali berada dalam *trade-off*, artinya peningkatan satu metrik dapat mengurangi yang lainnya. Untuk mengatasi hal tersebut, *F1-score* dapat digunakan sebagai metrik evaluasi, yang merupakan rata-rata harmonis dari *precision* dan *recall*. *F1-score* memberikan gambaran yang lebih seimbang tentang kinerja model, *F1-score* sangat berguna, terutama ketika menghadapi data yang tidak seimbang. *F1-score* dihitung dengan rumus:

$$F1 - Score = 2x \frac{Precision \times Recall}{Precision + Recall}$$

F1-score sangat berguna ketika diperlukan keseimbangan antara *precision* dan *recall*, terutama dalam situasi di mana kesalahan prediksi baik yang positif maupun negatif dapat memiliki dampak yang signifikan terhadap hasil analisis atau keputusan bisnis.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

2.9 Python

Python merupakan salah satu bahasa pemrograman yang paling populer dan banyak digunakan dalam *data science* dan *machine learning*, berkat fleksibilitasnya, kemudahan penggunaan, dan ekosistem pustaka (*library*) yang sangat kuat. Salah satu alasan mengapa *Python* sangat disukai oleh para *data scientist* dan peneliti adalah karena sintaksisnya yang mudah dipahami, sehingga memungkinkan pengguna, baik pemula maupun profesional, untuk lebih fokus pada analisis data dan penerapan algoritma daripada harus khawatir tentang kompleksitas bahasa pemrograman itu sendiri. Dalam konteks penelitian ini, *Python* memainkan peran yang sangat penting dalam beberapa tahap, mulai dari *preprocessing* data hingga pelatihan dan evaluasi model *machine learning*, khususnya dengan menggunakan algoritma *Gradient Boosting* untuk mengklasifikasikan loyalitas pelanggan di platform *e-commerce* Shopee.[17]

Selain itu, *Python* juga sangat berguna dalam **evaluasi model**. Setelah model dilatih menggunakan data pelatihan, langkah selanjutnya adalah mengevaluasi kinerja model tersebut dengan menggunakan data pengujian. *Python* menyediakan pustaka yang memudahkan peneliti dalam melakukan evaluasi ini. Salah satu pustaka utama yang digunakan adalah *scikit-learn*, yang tidak hanya memberikan berbagai algoritma untuk klasifikasi tetapi juga berbagai metrik evaluasi yang berguna, seperti *confusion matrix*, *classification report*, dan *ROC curve*. Dengan *confusion matrix*, peneliti dapat mengetahui jumlah *true positives*, *true negatives*, *false positives*, dan *false negatives* yang dihasilkan oleh model, yang sangat penting dalam menilai kesalahan prediksi yang dilakukan oleh model. Selain itu, menggunakan *classification report* memungkinkan perhitungan metrik penting lainnya, seperti *precision*, *recall*, dan *F1-score*, yang sangat berguna untuk mengevaluasi kinerja model dalam mengklasifikasikan pelanggan yang loyal dan tidak loyal.

Evaluasi yang dilakukan dengan menggunakan *Python* juga dapat diperluas dengan visualisasi hasil klasifikasi. Pustaka Python seperti *matplotlib* dan *seaborn* memungkinkan peneliti untuk menggambarkan hasil analisis data

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

dan evaluasi model dalam bentuk grafik yang mudah dipahami. Visualisasi ini dapat berupa grafik *confusion matrix* atau grafik perbandingan antara model yang satu dengan lainnya, atau visualisasi distribusi nilai yang diprediksi oleh model. Misalnya, dengan menggunakan *seaborn*, peneliti dapat dengan jelas melihat bagian-bagian mana dari *confusion matrix* yang mengandung kesalahan prediksi terbesar, sehingga langkah-langkah perbaikan dapat diambil untuk meningkatkan kinerja model.

Python juga memberikan kemudahan dalam **penanganan data besar**. Dalam dunia *e-commerce* seperti Shopee, volume data yang dihasilkan sangat besar, dan *Python* memiliki berbagai pustaka yang dapat menangani data dalam jumlah besar dengan efisien. Pustaka seperti *Dask* dan *PySpark* memungkinkan pengolahan data besar dengan mendistribusikan beban kerja ke berbagai core prosesor atau mesin, yang memungkinkan analisis yang lebih cepat dan *skalabel*. Penggunaan pustaka-pustaka ini sangat bermanfaat ketika *dataset* yang digunakan dalam penelitian ini mencakup banyak transaksi pelanggan dan interaksi yang membutuhkan pemrosesan data dalam jumlah besar dan waktu yang relatif singkat.

2.10 Flowchart


Flowchart adalah salah satu alat yang sangat penting dalam menggambarkan secara visual urutan langkah-langkah yang terjadi dalam suatu proses. Dalam konteks penelitian yang berfokus pada klasifikasi loyalitas pelanggan menggunakan algoritma *Gradient Boosting*, *Flowchart* berfungsi sebagai representasi visual dari keseluruhan alur kerja yang terlibat, mulai dari pengumpulan data, *preprocessing* data, pelatihan model, hingga evaluasi hasil klasifikasi. Dengan menggunakan *Flowchart*, langkah-langkah yang kompleks dan saling terkait dalam proses penelitian ini dapat disajikan secara sistematis dan mudah dipahami, baik oleh peneliti itu sendiri maupun pihak lain yang berkepentingan.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

2.2 Penelitian Terdahulu Yang Relevan

Tabel 2.1 Penelitian Terdahulu



No	Nama Penerbit & Tahun	Judul	Metode	Hasil
1	Sholeha, S. H., Faid, M., & Yaqin, M. A. (2024).	Prediksi Perpindahan Pelanggan Pada Toko <i>Online</i> Menggunakan Metode Tree- Based Gradient Boosted Models.	Metode penelitian ini melibatkan pengumpulan data, praproses data, pemilihan dan pelatihan model, serta evaluasi model. Pendekatan yang digunakan adalah data mining dengan fokus pada algoritma pembelajaran mesin, khususnya metode <i>tree-based gradient boosted models</i> , yang mencakup model <i>XGBoost</i> , <i>LightGBM</i> , dan <i>CatBoost</i> . Penelitian ini menggunakan	Hasil penelitian menunjukkan bahwa model XGBoost mencapai akurasi tertinggi dalam memprediksi churn pelanggan, dengan nilai ROC curve sebesar 0.66 dan akurasi 0.80032 . Analisis pentingnya fitur menunjukkan bahwa variabel gender memiliki pengaruh yang signifikan terhadap kinerja model. Penelitian ini memberikan kontribusi dalam meningkatkan layanan pelanggan, mengurangi churn,

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



berbagai pustaka dan pada akhirnya

seperti `pandas`, meningkatkan

`numpy`, dan profitabilitas

`matplotlib` untuk perusahaan di

analisis data dan sektor toko *online*.

pelatihan model.

2. Mohamed M. Abbassy. (2023). *Using Machine Learning Technique for Analytical Customer Loyalty.* Metode penelitian ini membangun model prediksi kepuasan pelanggan menggunakan machine learning dengan langkah-langkah berikut: Memahami data dari ulasan pelanggan di *e-commerce*, Memproses data melalui tokenisasi, pembersihan, dan penghapusan kata henti, Membangun model menggunakan algoritma seperti Random Forest dan SVM, Mengevaluasi model dengan metode *sampling stratified 10-fold*, Menganalisis hasil untuk menentukan efektivitas model.
 - Algoritma Gradient Boosing mencapai akurasi tertinggi 88%.
 - Akurasi kepuasan mengukur efektivitas model, tidak mempertimbangkan varians kelas.
 - Pelanggan yang dipertahankan memiliki rata-rata 18 hari aktif.
 - Pelanggan yang bergejolak memiliki rata-rata 1 hari aktif.
 - Ketidakseimbangan kelas diatasi menggunakan metode pengambilan sampel.
 - Berbagai

Protected by PDF Anti-Copy Free


(Upgrade to Pro Version to Remove the Watermark)



- pengklasifikasi dievaluasi untuk prediksi kepuasan pelanggan.
3. Nasyuli, L., Lubis, I., & Elhanafi, A. M. (2023). Penerapan Metode yang Hasil dari Model Machine digunakan dalam penelitian ini & Elhanafi, Learning penelitian ini menunjukkan A. M. (2023). Algoritma adalah **data** bahwa penerapan *Gradient* **mining** untuk algoritma *Gradient Boosting* dan memprediksi harga *Boosting* dan Linear Regression mobil bekas, *linear regression* Regression khususnya mobil berhasil Melakukan roda empat. menghasilkan Prediksi Harga Penelitian ini prediksi harga Kendaraan menggunakan dua mobil bekas yang Bekas. algoritma untuk akurat. Dengan menghasilkan menggunakan prediksi, yaitu kedua metode *Gradient Boosting* tersebut, penelitian *algorithm* dan ini memberikan *linear regression* kontribusi dalam *algorithm*. Kedua mengurangi algoritma tersebut ketidakpastian diterapkan untuk harga yang sering menganalisis data dihadapi oleh harga mobil bekas pembeli mobil yang bervariasi bekas, serta dan mempermudah membandingkan transaksi jual beli hasil prediksi dari kendaraan roda masing-masing empat di kota

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

			metode.	Medan.
4	Xujie Qin (2023).	Research loyalty prediction <i>commerce</i> customer based on data mining.	 <p>Metode penelitian ini mencakup pengumpulan data dari platform <i>e-commerce</i>, pembersihan data dengan algoritma, analisis klasifikasi loyalitas pelanggan, modeling menggunakan algoritma XGBoost, evaluasi model, dan analisis hasil untuk memahami loyalitas pelanggan.</p>	<p>Model XGBoost yang dioptimalkan menunjukkan akurasi tinggi dalam prediksi. Tingkat Akurasi, Presisi, dan Penarikan adalah 0,9213, 0,9412, dan 0,8712. Nilai F dari model adalah 0,9048. Model ini secara efektif menangani masalah data yang hilang dan tidak seimbang</p>
5	Purnama, J., Hikmawati, N. K., & Rahayu, S. (2024).	Analisis Algoritma Klasifikasi Untuk Mengidentifikasi Potensi Risiko Kesehatan Ibu Hamil.	<p>Penelitian ini menggunakan teknik <i>machine learning</i> untuk mengklasifikasikan potensi risiko kesehatan pada ibu hamil, dengan menganalisis dataset kesehatan</p>	<p>Hasil penelitian menunjukkan bahwa algoritma <i>Random Forest</i> memberikan kinerja terbaik dengan akurasi sebesar 82,15% dalam mengklasifikasikan</p>

Protected by PDF Anti-Copy Free

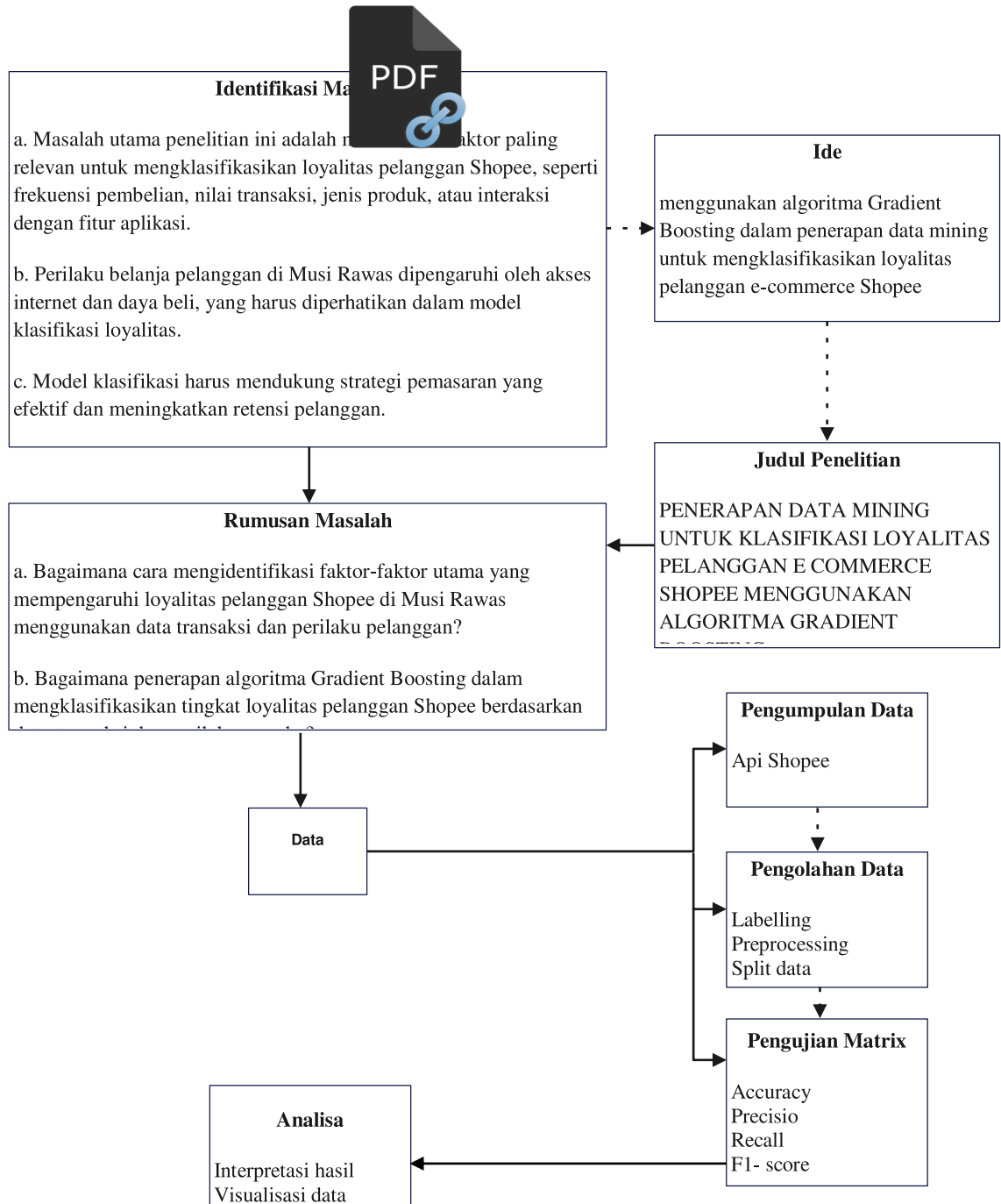
(Upgrade to Pro Version to Remove the Watermark)



maternal. Beberapa risiko kesehatan algoritma yang maternal pada tiga diterapkan untuk tingkat: tinggi, klasifikasi data sedang, dan rendah. adalah **Random** Temuan ini **Forest**, **Extra** menunjukkan **Trees**, **Extreme** bahwa model yang **Gradient** dibuat dapat **Boosting**, **Decision** mengidentifikasi **Tree**, dan **Light** pola dan hubungan **Gradient Boosting** kompleks antara **Machine**. Setiap fitur yang relevan algoritma diuji dan dalam klasifikasi dievaluasi untuk risiko kesehatan ibu melihat kinerja dan hamil, yang dapat akurasi dalam membantu tenaga mengidentifikasi medis dalam risiko kesehatan pengambilan keputusan yang maternal. lebih efektif dan memberikan wawasan lebih awal kepada ibu hamil mengenai kondisi kesehatan mereka.

Protected by PDF Anti-Copy Free

2.3 (Upgrade to Pro Version to Remove the Watermark) Kerangka Berpikir



Gambar 2.1 Kerangka Berpikir

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Kerangka berpikir merupakan dasar pemikiran dalam menentukan alur dari penelitian. Dengan adanya kerangka berpikir maka penelitian dapat tersusun rapi dan mudah dipahami. Penelitian ini diawali dengan identifikasi masalah mengklasifikasi loyalitas pelanggan pada platform *e-commerce* Shopee, yang diikuti dengan rumusan masalah yang jelas, menghasilkan judul penelitian "Penerapan Data mining untuk Klasifikasi Loyalitas Pelanggan *E-commerce* Shopee Menggunakan Algoritma *Gradient Boosting*". Penelitian ini melibatkan pengumpulan data melalui *API* Shopee untuk mendapatkan informasi yang diperlukan, yang kemudian diproses dengan melakukan pelabelan, *preprocessing*, dan pemisahan data untuk mempersiapkannya dalam analisis. Selanjutnya, pengujian dilakukan melalui penghitungan matriks evaluasi, termasuk akurasi, *precision*, *recall*, dan *F1-score* untuk menilai kinerja model. Hasil dari proses tersebut akan dianalisis lebih lanjut dengan menampilkan interpretasi data yang diperoleh dan visualisasi yang jelas untuk memberikan wawasan mendalam mengenai loyalitas pelanggan di Shopee.



3.1 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah pendekatan kuantitatif dengan desain eksploratori dan deskriptif, yang bertujuan untuk memahami, menggali, dan mengklasifikasikan loyalitas pelanggan Shopee berdasarkan karakteristik pelanggan dan data transaksi mereka. Pendekatan kuantitatif dipilih karena memberikan kemampuan untuk menganalisis hubungan antara berbagai variabel yang dapat diukur secara numerik.

3.2 Metode Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan dua metode utama: **pengumpulan data primer** dan **pengumpulan data sekunder**. Data primer akan diperoleh melalui akses langsung terhadap data pelanggan Shopee yang bertransaksi di wilayah Musi Rawas, Sumatera Selatan, sedangkan data sekunder diperoleh dari sumber-sumber terkait yang mendukung analisis dalam penelitian ini.

3.2.1 Data Primer

Data primer akan diperoleh melalui database internal Shopee yang mencakup dua jenis data utama: **data karakteristik pelanggan** dan **data transaksi pelanggan**. Data karakteristik pelanggan mencakup informasi seperti usia, jenis kelamin, dan lokasi pelanggan. Sementara itu, data transaksi mencakup informasi tentang produk yang dibeli, nilai transaksi, frekuensi pembelian, serta metode pembayaran.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3.2.2 Data Sekunder

Data sekunder diperoleh dari berbagai literatur, jurnal penelitian, laporan industri, dan sumber-sumber yang membahas mengenai perilaku pelanggan, *e-commerce*, dan loyalitas pelanggan dalam konteks *e-commerce*.

3.3 Metode Analisa

Dalam penelitian ini, analisis data dilakukan dengan pendekatan *machine learning* menggunakan algoritma *Gradient Boosting*. *Dataset* yang dianalisis berjumlah 2000 dengan pemilihan berdasarkan kriteria representativitas untuk mencerminkan populasi pelanggan. Proses seleksi ini bertujuan meningkatkan kualitas analisis dan menghindari overfitting [18]. Dari 2000 data yang terpilih, 80% digunakan untuk pelatihan dan 20% untuk pengujian, yang terdiri dari dua kategori: loyal dan tidak loyal. Proses analisis ini melibatkan beberapa langkah penting, mulai dari pra-pemrosesan data hingga penerapan algoritma untuk memprediksi loyalitas pelanggan. Berikut adalah tahapan metode analisis yang digunakan dalam penelitian ini:

3.3.1 Pra-Pemrosesan Data

a. Pembersihan Data

Pada tahap awal, data yang telah dikumpulkan akan dibersihkan dari duplikasi, data yang hilang (*missing values*), dan kesalahan yang mungkin ada dalam data.

b. Transformasi Data

Data yang telah dibersihkan akan diubah atau diproses dalam format yang sesuai dengan algoritma yang digunakan.

c. Normalisasi Data

Normalisasi sangat penting dalam *machine learning*, terutama ketika data yang digunakan memiliki rentang nilai yang berbeda-beda.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3.3.2 Penerapan Algoritma *Gradient Boosting*

a. Pembagian Data

Data yang telah diproses dibagi menjadi dua set, yaitu data latih (training data) dan data uji (test data). Data latih akan digunakan untuk melatih model, sedangkan data uji digunakan untuk menguji akurasi model yang telah dilatih.

b. Model *Gradient Boosting*

Algoritma *Gradient Boosting* akan diterapkan pada data untuk memprediksi loyalitas pelanggan berdasarkan variabel independen yang telah ditentukan, yaitu karakteristik pelanggan (X1) dan data transaksi (X2).

c. Pengujian Model

Setelah model dilatih, dilakukan evaluasi terhadap hasil prediksi yang dihasilkan menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai kinerja model dalam mengklasifikasikan loyalitas pelanggan.

3.3.3 Analisis Hasil

a. Interpretasi Hasil

Hasil yang diperoleh dari model *Gradient Boosting* akan dianalisis untuk melihat faktor-faktor mana saja yang paling berpengaruh terhadap loyalitas pelanggan. Faktor-faktor ini akan dijadikan dasar dalam merumuskan rekomendasi bagi pihak Shopee dalam meningkatkan loyalitas pelanggan mereka.

b. Visualisasi Data

Untuk mempermudah pemahaman, hasil analisis akan divisualisasikan dalam bentuk grafik, tabel, dan diagram yang menggambarkan hubungan antar variabel serta hasil prediksi dari model.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3.4 Tempat dan Waktu Penelitian

3.4.1 Tempat



Penelitian ini dilakukan di wilayah Musi Rawas, yang terletak di Sumatera Selatan. Musi Rawas dipilih sebagai lokasi penelitian karena merupakan daerah dengan populasi yang cukup besar dan cukup representatif dalam menggambarkan perilaku pelanggan *e-commerce* di luar kota besar.

3.4.2 Waktu

Waktu Penelitian: Penelitian ini dilakukan selama periode 5 bulan, dimulai pada bulan September 2024 hingga Januari 2025. Dalam rentang waktu tersebut, peneliti akan mengumpulkan data, melakukan analisis, serta menyusun laporan hasil penelitian. Pengumpulan data dilakukan pada bulan pertama, sementara analisis data dilakukan pada bulan kedua hingga kelima. Bulan terakhir digunakan untuk menyusun hasil penelitian dan menyelesaikan laporan.

Tabel 3.1 Jadwal Penelitian

No	Jenis Kegiatan	Waktu Kegiatan																			
		September 2024				Oktober 2024				November 2024				Desember 2024				Januari 2025			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengajuan Judul	■	■	■	■																
2	Pengumpulan Data					■	■	■	■												
3	Penulisan Proposal									■	■	■	■								
4	Bimbingan Proposal													■	■	■	■				
5	Ujian Proposal																	■	■	■	■
6	Revisi Ujian Proposal																				
7	Pengolahan dan Pengujian Data																				
8	Bimbingan Skripsi																				
9	Ujian Skripsi																				

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3.5 Alat dan Bahan

3.5.1 Alat

Dalam melakukan penelitian tentu saja diperlukan alat untuk mendukung proses penelitian. Berikut adalah alat tersebut sebagai berikut:



Tabel 3.2 Alat

No	Alat
1	Laptop Acer Aspire
2	Python (IDE: Jupyter Notebook)
3	Software Pengolah Data (Visual Studio Code)
4	API Shopee (untuk mengambil data)
5	Software microsoft word
6	Printer Epson L5190

3.5.2 Bahan

Bahan yang diperlukan dalam proses penelitian sebagai berikut:

Tabel 3.3 Bahan

No	Bahan
1	Data transaksi pelanggan Shopee (melalui API Shopee)
2	Dataset yang sudah diproses dan di-labeling
3	Skrip Python untuk preprocessing data, implementasi model, dan evaluasi

3.6 Metode Pengujian dan Pengolahan Data

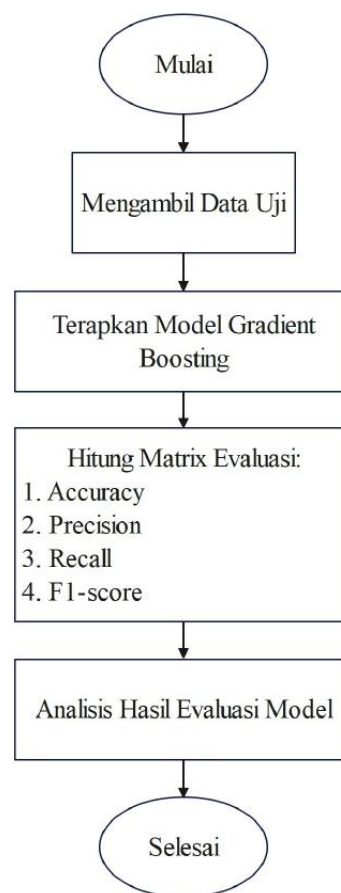
Pengujian dan pengolahan data dalam penelitian ini dilakukan dengan serangkaian langkah yang mencakup pra-pemrosesan data, pelatihan model dengan algoritma *Gradient Boosting*, serta evaluasi dan interpretasi hasil prediksi.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3.6.1 Pengujian Data

Metode pengujian dalam penelitian ini digunakan untuk mengevaluasi performa klasifikasi yang telah dilatih, dengan tujuan untuk mengukur sejauh mana model dapat mengklasifikasikan loyalitas pelanggan Shopee secara akurat. Proses pengujian ini sangat penting untuk mengetahui kualitas dan keandalan model dalam memberikan prediksi yang benar.



Gambar 3.1 Flowchart Pengujian Model

a. Mengambil Data Uji

Pengujian dimulai dengan mengambil data uji (test data) yang sebelumnya telah dipisahkan dari data pelatihan. Data uji ini tidak digunakan selama pelatihan model sehingga bisa memberikan

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

gambaran yang lebih akurat mengenai kinerja model pada data yang belum pernah dilihat sebelumnya.

b. Terapkan Model *Gradient Boosting*

Pada langkah ini, model yang telah dilatih pada data pelatihan diterapkan untuk melakukan prediksi terhadap data uji. Model klasifikasi *Gradient Boosting* akan memberikan prediksi apakah pelanggan termasuk dalam kategori loyal atau tidak loyal.

c. Hitung Metrik Evaluasi

Selanjutnya, hasil prediksi dari model dibandingkan dengan label yang sebenarnya (*ground truth*) dalam data uji. Berdasarkan perbandingan ini, metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* dihitung. Metrik ini digunakan untuk menilai seberapa baik model dalam mengklasifikasikan data.

- a) **Accuracy:** Persentase prediksi yang benar dari seluruh prediksi yang dilakukan.
- b) **Precision:** Mengukur akurasi prediksi positif yang dilakukan oleh model.
- c) **Recall:** Mengukur kemampuan model dalam mengidentifikasi semua contoh positif yang sebenarnya.
- d) **F1-score:** Rata-rata harmonik antara precision dan recall, memberikan gambaran seimbang antara keduanya.

d. Analisis Hasil Evaluasi Model

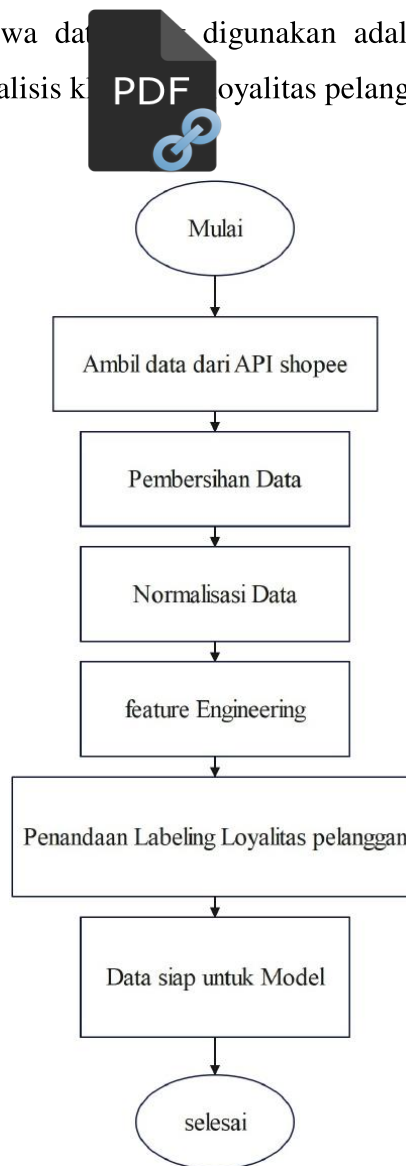
Setelah metrik dihitung, hasil evaluasi akan dianalisis untuk mengetahui kelebihan dan kekurangan model. Apakah model sudah cukup baik dalam memprediksi loyalitas pelanggan atau perlu dilakukan perbaikan.

3.6.2 Pengolahan Data

Metode pengolahan data dalam penelitian ini adalah serangkaian langkah-langkah yang dilakukan untuk mempersiapkan dan membersihkan data sebelum digunakan untuk pelatihan model *machine learning*. Data

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)
 yang diperoleh dari Shopee perlu diolah dengan hati-hati untuk memastikan bahwa data yang digunakan adalah data yang bersih dan relevan untuk analisis ke-loyalitas pelanggan.



Gambar 3.2 Flowchart Pengolahan Data

a. Ambil Data dari API Shopee

Proses pengolahan data dimulai dengan mengambil data transaksi pelanggan Shopee menggunakan API yang telah disiapkan. Data ini berisi informasi tentang pelanggan, transaksi yang dilakukan, produk yang dibeli, dan interaksi dengan platform.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

b. Pembersihan Data

Setelah data siap, langkah pertama dalam pengolahan adalah pembersihan data. Pada tahap ini, peneliti akan menghapus duplikasi data, mengisi nilai yang hilang (*missing values*), dan memperbaiki inkonsistensi dalam data. Pembersihan data penting untuk memastikan bahwa data yang digunakan dalam model tidak mengandung kesalahan yang dapat mempengaruhi hasil analisis.

c. Normalisasi Data

Setelah pembersihan, data perlu dinormalisasi untuk memastikan bahwa setiap fitur berada pada skala yang seragam. Normalisasi sangat penting dalam *machine learning*, terutama ketika data yang digunakan memiliki rentang nilai yang berbeda-beda. Misalnya, frekuensi pembelian bisa memiliki skala yang jauh lebih besar dibandingkan dengan nilai transaksi. Proses normalisasi akan membuat fitur-fitur ini lebih setara dalam kontribusinya terhadap model.

d. Feature Engineering

Pada tahap ini, dilakukan penambahan atau modifikasi fitur yang dapat membantu model dalam meningkatkan performa klasifikasi. *Feature engineering* bisa melibatkan pembuatan fitur baru berdasarkan data yang ada, seperti menggabungkan informasi frekuensi pembelian dengan nilai transaksi untuk membuat fitur "total pembelian". Langkah ini bertujuan untuk meningkatkan kualitas input data agar model dapat mempelajari pola dengan lebih baik.

e. Penandaan *Labeling* Loyalitas Pelanggan

Setelah menyiapkan data, langkah selanjutnya adalah proses penandaan untuk mengklasifikasikan pelanggan sebagai loyal atau tidak loyal. Penandaan label loyalitas pelanggan dilakukan dengan cara yang sistematis, di mana pelanggan akan diklasifikasikan sebagai "Loyal" jika mereka memenuhi setidaknya salah satu dari kriteria yang telah ditentukan. Kriteria tersebut meliputi: jumlah pembelian

Protected by PDF Anti-Copy Free

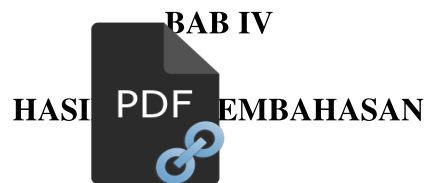
(Upgrade to Pro Version to Remove the Watermark)

yang dilakukan mencapai atau lebih dari 4 kali, total pengeluaran yang dikeluarkan oleh pelanggan mencapai Rp177.000 atau lebih, serta rata-rata jumlah pembelian yang dilakukan per transaksi tidak kurang dari Rp62.000. Untuk memudahkan dalam memahami dan memberikan gambaran yang lebih jelas mengenai pengeluaran pelanggan, ambang batas total pengeluaran dan rata-rata pembayaran tersebut dikonversi dari nilai normalisasi ke dalam bentuk nilai rupiah. Dengan demikian, proses ini tidak hanya membantu dalam penentuan loyalitas pelanggan, tetapi juga memberikan pemahaman yang lebih mendalam tentang perilaku pengeluaran mereka. Di sisi lain, pelanggan yang tidak memenuhi salah satu dari kriteria yang telah ditetapkan tersebut akan diberikan label "Tidak Loyal", yang menunjukkan bahwa mereka tidak memenuhi standar yang dianggap sebagai loyalitas dalam konteks ini.

f. Siap untuk Model Training

Setelah data diproses, dibersihkan, dan dilabeli, data siap digunakan untuk melatih model. Data yang sudah diproses dengan baik akan memberikan hasil yang lebih akurat saat diterapkan pada algoritma *machine learning* seperti *Gradient Boosting*.

Metode pengolahan data yang sistematis dan terstruktur sangat penting untuk memastikan bahwa data yang digunakan dalam model *machine learning* relevan dan dapat memberikan hasil yang optimal. Dalam penelitian ini, pengolahan data dilakukan dengan hati-hati untuk menghindari kesalahan yang dapat mengurangi akurasi model.



4.1. Gambaran Umum

Shopee merupakan salah satu platform e-commerce terkemuka di kawasan Asia Tenggara dan Taiwan, yang telah berhasil menarik perhatian jutaan pengguna sejak peluncurannya pada tahun 2015. Berdasarkan data dari iPrice, pada kuartal ketiga tahun 2022, Shopee mencatat lebih dari 200 juta pengunjung bulanan di situsnya di Indonesia, menjadikannya salah satu platform perdagangan digital dengan pengunjung terbanyak. Keberhasilan ini didorong oleh berbagai fitur yang menawarkan kemudahan dan keamanan dalam berbelanja, serta strategi promosi yang agresif, seperti diskon dan flash sale, yang berkontribusi pada peningkatan keterlibatan pengguna [19].

Algoritma pembelajaran mesin seperti gradient boosting terbukti efektif dalam meramalkan perilaku konsumen berdasarkan data transaksi serta preferensi belanja mereka [20]. Dalam hal ini, analisis dilakukan dengan menggunakan dataset yang mencakup informasi pelanggan dan transaksi pada platform Shopee. Data tersebut dianalisis dengan menerapkan algoritma gradient boosting untuk mengklasifikasikan tingkat loyalitas pelanggan menjadi dua kategori, yaitu loyal dan tidak loyal. Pendekatan ini memanfaatkan keunggulan gradient boosting dalam mengelola dataset yang kompleks dan bervariasi, sehingga memungkinkan pengidentifikasian pola-pola penting dalam perilaku konsumen.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4.2. Hasil

4.2.1. Pengumpulan Data

Pengumpulan data menggunakan Python dan data transaksi di Shopee bertujuan untuk menganalisis perilaku pengguna dan mengoptimalkan layanan yang ditawarkan. Proses pengambilan data ini dilakukan melalui API Shopee, yang memerlukan kredensial seperti access token untuk mengautentikasi permintaan data. Berikut adalah ilustrasi atau gambar dari kode proses pengambilan data menggunakan API Shopee:

```
MULAI
# Import library yang diperlukan
import requests
import csv
import time
from datetime import datetime
from dotenv import load_dotenv
import os

# Memuat file .env untuk mengamankan data sensitif
load_dotenv()

# Ambil data sensitif dari file .env
PARTNER_ID = os.getenv("200****")
SHOP_ID = os.getenv("13****")
SECRET_KEY = os.getenv("9da4139*****")
ACCESS_TOKEN =
os.getenv("5977717463*****")

# URL dasar untuk API Shopee
BASE_URL_CUSTOMERS =
"https://partner.shopeemobile.com/api/v2/customer/get"
BASE_URL_ORDERS =
"https://partner.shopeemobile.com/api/v2/order/get"

# Fungsi untuk membuat signature (Shopee biasanya
memerlukan tanda tangan digital)
def generate_signature(base_url, params):
    import hashlib
    import hmac
    sorted_params = sorted(params.items())
    encoded_params = "&".join(f"{k}={v}" for k, v
in sorted_params)
    base_string = f"{base_url}{encoded_params}"
    signature = hmac.new(
        bytes(SECRET_KEY, "utf-8"),
        bytes(base_string, "utf-8"),
        hashlib.sha256
    ).hexdigest()
    return signature
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```

# Fungsi untuk mengambil data pelanggan
def fetch_order_data(start_date, end_date):
    print("Mengambil data pelanggan...")
    customer_data = []
    for page in range(1, 31):
        params = {
            "partner_id": PARTNER_ID,
            "shop_id": SHOP_ID,
            "timestamp":
                int(datetime.now().timestamp()),
            "time_from":
                int(start_date.timestamp()),
            "time_to": int(end_date.timestamp()),
            "page": page,
            "limit": 2000
        }
        params["signature"] =
            generate_signature(BASE_URL_CUSTOMERS, params)

        response = requests.get(BASE_URL_CUSTOMERS,
            params=params, headers={"Authorization":
                ACCESS_TOKEN})
        if response.status_code == 2000:
            result = response.json()
            if result["customers"]:
                customer_data.extend(result["customers"])
        else:
            break
    else:
        print(f"Error saat mengambil data
            pelanggan: {response.status_code},
            {response.text}")
        break
    time.sleep(1) # Jeda untuk menghindari
    rate limit
    return customer_data

# Fungsi untuk mengambil data transaksi
def fetch_order_data(start_date, end_date):
    print("Mengambil data transaksi...")
    order_data = []
    for page in range(1, 31):
        params = {
            "partner_id": PARTNER_ID,
            "shop_id": SHOP_ID,
            "timestamp":
                int(datetime.now().timestamp()),
            "time_from":
                int(start_date.timestamp()),
            "time_to": int(end_date.timestamp()),
            "page": page,
            "limit": 2000
        }
        params["signature"] =

```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```

generate_signature(BASE_URL_ORDERS, params)

response = requests.get(BASE_URL_ORDERS,
params=params, headers={"Authorization":
ACCESS_TOKEN})
if response.status_code == 2000:
    result = response.json()
    if result["orders"]:
        order_data.extend(result["orders"])
    else:
        break
else:
    print(f"Error saat mengambil data
transaksi: {response.status_code},
{response.text}")
    break
    time.sleep(1) # Jeda untuk menghindari
rate limit
return order_data

# Fungsi untuk menyimpan data ke CSV
def save_to_csv(data, filename, fieldnames):
    print(f"Menyimpan data ke {filename}...")
    with open(filename, mode="w", newline="",
encoding="utf-8") as file:
        writer = csv.DictWriter(file,
fieldnames=fieldnames)
        writer.writeheader()
        writer.writerows(data)
    print(f"Data berhasil disimpan ke {filename}.")

# Rentang waktu untuk data Januari 2024 hingga
Desember 2024
start_date = datetime(2024, 1, 1)
end_date = datetime(2024, 12, 31)

# Ambil data pelanggan dan transaksi
customer = fetch_customer_data(start_date,
end_date)
order = fetch_order_data(start_date, end_date)

# Simpan data ke CSV
if customer:
    save_to_csv(
        customer,
        "data_pelanggan_2024.csv",
        fieldnames=["customer_id", "name", "age",
"gender", "email", "phone", "address"]
    )

if order:
    save_to_csv(
        order,
        "data_transaksi_2024.csv",
        fieldnames=["transaction_id",
"customer_id", "customer_name", "product_name",

```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```

"category", "price", "quantity", "total_price",
"payment_method"]
)
print("Pro PDF sai.")
SELESAI

```

Gambar 4.1 Pseudocode Pengambilan Data

Pengumpulan informasi dilakukan dalam periode Januari 2024 hingga Desember 2024 dengan memanfaatkan Shopee API, yang terdiri dari dua endpoint utama, yaitu customer/get untuk memperoleh data pelanggan dan order/get untuk mendapatkan data transaksi. Data diambil secara massal dengan menerapkan pagination, dengan batasan maksimal 2000 data per permintaan, adapun hasilnya adalah sebagai berikut :

a. Data Pelanggan

	Customer ID	Nama Customer	Usia	Jenis Kelamin	Email	Phone	Alamat
0	855409	Hafshah Wibowo	19	Perempuan	hafshahwibowo929@gmail.com	8273735110	Jl. Desa Muara Kelingi No. 1, Kec. Muara Kelin...
1	617884	Taufik Sudiati	16	Perempuan	taufiksudiati275@gmail.com	8836557020	Jl. A. Yani No. 92, Kec. Muara Beliti, Kab. Mu...
2	336002	Wadi Permadi	26	Perempuan	wadipermadi321@gmail.com	8387244347	Jl. Desa Purwodadi No. 48, Kec. Purwodadi, Kab...
3	555351	Bajragin Yuliarti	35	Laki-Laki	bajraginyuliarti195@gmail.com	8784425430	Jl. Desa Muara Kelingi No. 23, Kec. Muara Keli...
4	742947	Kariman Nashiruddin	40	Laki-Laki	karimannashiruddin721@gmail.com	8721553149	Jl. Desa Karang Anyar No. 105, Kec. Megang Sak...

Gambar 4.2 Hasil Pengambilan Data Pelanggan

b. Data Transaksi

	Transaction ID	Customer ID	Nama Customer	Nama Produk	Kategori	Harga	Jumlah Pembelian	Total Harga	Metode Pembayaran
0	3290956	855409	Hafshah Wibowo	Headphone	Elektronik	1342810	4	5371240	COD
1	8512438	617884	Taufik Sudiati	T-Shirt	Fashion	76557	4	306228	COD
2	8143910	336002	Wadi Permadi	Lipstick	Kecantikan	194619	4	778476	Transfer Bank
3	6811827	555351	Bajragin Yuliarti	Smartphone	Elektronik	2839961	1	2839961	Seabank
4	1873087	742947	Kariman Nashiruddin	Jacket	Fashion	599700	1	599700	COD

Gambar 4.3 Hasil Pengambilan Data Transaksi

4.2.2. Preprocessing

Preprocessing merupakan langkah awal dalam menyiapkan data sebelum diolah atau dianalisis. Tujuan dari proses ini adalah untuk

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

membersihkan data agar siap digunakan oleh algoritma. Berikut adalah beberapa tahap dalam proses ini.

4.2.2.1. Pembersihan Data

Proses pembersihan data bertujuan untuk memperbaiki mutu data dengan cara menghapus elemen-elemen yang tidak diperlukan, menyelaraskan format, serta memastikan tidak adanya pengulangan data yang dapat berdampak pada analisis. Adapun hasil pembersihan data sebagai berikut :

Customer ID	Nama Customer	Usia	Jenis Kelamin	Email	Phone	Alamat	Transaction ID	Nama Produk	Kategori	Harga	Jumlah Pembelian	Total Harga	Metode Pembayaran	
0	855409	Hafshah Wibowo	19	Perempuan	hafshahwibowo929@gmail.com	8273735110	Jl. Desa Muara Keilingi No. 1, Kec. Muara Kelin...	3290956	Headphone	Elektronik	1342810	4	5371240	COD
1	617884	Taufik Sudiati	16	Perempuan	taufiksudiati275@gmail.com	8836557020	Jl. A. Yani No. 92, Kec. Muara Beliti, Kab. Mu...	8512438	T-Shirt	Fashion	76557	4	306228	COD
2	336002	Wadi Permadi	26	Perempuan	wadipermadi321@gmail.com	8387244347	Jl. Desa Punvodadi No. 48, Kec. Punvodadi, Kab...	8143910	Lipstick	Kecantikan	194619	4	778476	Transfer Bank
3	555351	Bajragin Yulianti	35	Laki-Laki	bajraginyulianti195@gmail.com	8784425430	Jl. Desa Muara Keilingi No. 23, Kec. Muara Kelin...	6811827	Smartphone	Elektronik	2839961	1	2839961	Seabank
4	742947	Kariman Nashiruddin	40	Laki-Laki	karimannashiruddin721@gmail.com	8721553149	Jl. Desa Karang Anyar No. 105, Kec. Megang Sak...	1873087	Jacket	Fashion	599700	1	599700	COD

Gambar 4.4 Hasil Pembersihan Data

Berikut pseudocode untuk proses pembersihan data sebagai berikut:

```
# Menghapus kolom duplikat
columns_to_drop = ['Customer ID.1', 'Nama Customer.1']
data_cleaned = data.drop(columns=columns_to_drop, axis=1)

# Menyelaraskan format data: ubah kolom 'Phone' menjadi string
if 'Phone' in data_cleaned.columns:
    data_cleaned['Phone'] = data_cleaned['Phone'].astype(str)

# Terakhir: Remove duplicate rows
data = data.drop_duplicates()
```

Gambar 4.5 Pseudocode Pembersihan Data

4.2.2.2. Normalisasi Data

Proses normalisasi data bertujuan untuk mengatur skala nilai dalam dataset agar berada dalam rentang tertentu, biasanya antara 0 dan 1. Tahapan ini sangat krusial, terutama ketika data

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

akan digunakan untuk analisis atau algoritma yang peka terhadap

skala, seperti *recommender system* atau *machine learning*. Berikut ini adalah

hasil normalisasi



Customer ID	Nama Customer	Usia	Jenis Kelamin	Phone	Alamat	Transaction ID	Nama Produk	Kategori	Harga	Jumlah Pembelian	Total Harga	Metode Pembayaran
0	855409 Hafshah Wilowo	19	Perempuan	hafshah...	Jl. Desa Muara Kelingi No. 1, Kec. Muara Keli...	3290956	Headphone	Elektronik	0.266505	4	0.225906	COD
1	617884 Taufik Sudiat	16	Perempuan	taufiksudiat275@gmail.com	8836557020	8512438	T-Shirt	Fashion	0.011328	4	0.012069	COD
2	336002 Wadi Permadi	26	Perempuan	wadipermadi321@gmail.com	8387244347	8143910	Lipstick	Kecantikan	0.035120	4	0.032007	Transfer Bank
3	555351 Bajragin Yuliant	35	Laki-Laki	bajraginyuliant195@gmail.com	8784425430	6811827	Smartphone	Elektronik	0.568214	1	0.119040	Seabank
4	742947 Kariman Nashruddin	40	Laki-Laki	karimannashruddin721@gmail.com	8721553149	1873087	Jacket	Fashion	0.116752	1	0.024459	COD

Gambar 4.6 Hasil Normalisasi Data

Berikut *pseudocode* normalisasi data sebagai berikut:

```
# Pilih kolom numerik yang akan dinormalisasi
columns_to_normalize = ['Harga', 'Total Harga']
# Lakukan normalisasi Min-Max
scaler = MinMaxScaler()
data_normalized = data.copy()
data_normalized[columns_to_normalize] =
scaler.fit_transform(data[columns to normalize])
```

Gambar 4.7 Pseudocode Normalisasi Data

4.2.2.3. Feature Engineering

Proses *feature engineering* bertujuan untuk menghasilkan atau memodifikasi fitur dalam kumpulan data sehingga menjadi lebih relevan, informatif, dan lebih mudah untuk diproses oleh algoritma analisis atau *machine learning*. Berikut ini adalah hasil *featur engineering*

:

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Customer ID	Nama Customer	Usia	Jenis Kelamin	Email	Phone	Alamat	Transaction ID	Nama Produk	Kategori	Harga	Jumlah Pembelian	Total Harga	Metode Pembayaran	Total Pembayaran Rata-rata	Kelompok Usia
0	855409	Hafshah Wibowo	19	1	hafshahwibowo@gmail.com	Jl. Desa Muara Kelinci No. 1, Kec. Muara Kelinci...	3290956	Headphone	0	0.266505	4	0.225906	0	0.056477	Remaja
1	617884	Taufik Sudiaty	16	1	taufikusudiaty@gmail.com	Jl. A. Yani No. 92, Kec. Muara Belit, Kab. Muar...	8512438	T-Shirt	1	0.011328	4	0.012069	0	0.003017	Anak-anak
2	336002	Wadi Permadi	26	1	wadipermadi@gmail.com	Jl. Desa Purwodadi No. 48, Kec. Purwodadi, Kab...	8143910	Lipstick	2	0.035120	4	0.032007	4	0.008002	Dewasa Muda
3	555351	Bajragin Yulianti	35	0	bajraginyulianti195@gmail.com	Jl. Desa Muara Kelinci No. 23, Kec. Muara Kelinci...	6811827	Smartphone	0	0.568214	1	0.119040	1	0.119040	Dewasa
4	742947	Kariman Nashruddin	40	0	karimannashruddin721@gmail.com	Jl. Desa Karang Anyar No. 105, Kec. Megang Sakti...	1873087	Jacket	1	0.116752	1	0.024459	0	0.024459	Dewasa

Gambar 4.8 Hasil *Feature Engineering*

Berikut pseudocode *feature engineering* sebagai berikut:

```
# ---- 1. Encoding Fitur Kategori ----
# Fitur kategori yang akan di-encode
kategori_fitur = ['Jenis Kelamin', 'Metode
Pembayaran', 'Kategori']
# Menggunakan Label Encoding
label_encoders = {}
for fitur in kategori_fitur:
    le = LabelEncoder()
    data[fitur] = le.fit_transform(data[fitur])
    label_encoders[fitur] = le # Menyimpan encoder
untuk digunakan kembali jika perlu
# Jika menggunakan One-Hot Encoding (pilihan
alternatif):
# data = pd.get_dummies(data,
columns=kategori_fitur, drop_first=True)

# ---- 2. Menambahkan Kolom Total Pembayaran Rata-
Rata ----
# Pastikan dataset memiliki kolom 'Total Pembayaran'
dan 'Jumlah Transaksi'
data['Total Pembayaran Rata-rata'] = data['Total
Harga'] / data['Jumlah Pembelian']
# ---- 3. Menambahkan Kolom Kelompok Usia ----
# Buat rentang usia dan labelnya
bins = [0, 17, 25, 35, 50, 100] # Rentang usia
labels = ['Anak-anak', 'Remaja', 'Dewasa Muda',
'Dewasa', 'Lansia']
data['Kelompok Usia'] = pd.cut(data['Usia'],
bins=bins, labels=labels, right=False)
# ---- 4. Pengelompokan Pelanggan ----
# Fitur yang digunakan untuk clustering
features = data[['Jumlah Pembelian', 'Total Harga']]
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```
# Inisialisasi model KMeans
kmeans = KMeans(n_clusters=3, random_state=42)
data['Kelompok Pelanggan'] =
kmeans.fit_predict(features)
```

Gambar 4.9 Pseudocode Feature Engineering

4.2.3. Labelling

Tahapan pelabelan digunakan untuk mengenali pelanggan setia berdasarkan kriteria tertentu, yaitu frekuensi pembelian, total belanja, dan rata-rata nilai transaksi. Batasan ditetapkan sebagai frekuensi pembelian ≥ 4 , total belanja pada kuantil ke-75% atau lebih, serta rata-rata nilai transaksi pada kuantil ke-75% atau lebih. Pelanggan akan diberi label sebagai setia (1) jika memenuhi salah satu dari kriteria tersebut. Berikut adalah hasil pemberian label loyalitas.


Nama Customer	Usia	Jenis Kelamin	Email	Phone	Alamat	Transaction ID	Nama Produk	Kategori	Harga	Jumlah Pembelian	Total Harga	Metode Pembayaran	Total Pembayaran Rata-rata	Kelompok Usia	Kelompok Pelanggan	Loyalitas
Hafshah Wibowo	19	1	hafshahwibowo929@gmail.com	8273735110	Jl. Desa Muara Kelingi No. 1, Kec. Muara Kelingi, Kab. Muara...	3290956	Headphone	0	0.266505	4	0.225906	0	0.056477	Remaja	0	1
Taufik Sudlati	16	1	taufiksudlati275@gmail.com	8836557020	Jl. A. Yani No. 92, Kec. Muara Belit, Kab. Mu...	8512438	T-Shirt	1	0.011328	4	0.012069	0	0.003017	Anak-anak	0	1
Wadi Permadi	26	1	wadipermadi321@gmail.com	8387244347	Jl. Desa Purwodadi No. 48, Kec. Purwodadi, Kab. ...	8143910	Lipstick	2	0.035120	4	0.032007	4	0.008002	Devasa Muda	0	1
Bajragin Yuliani	35	0	bajraginyuliani195@gmail.com	8784425430	Jl. Desa Muara Kelingi No. 23, Kec. Muara Keli...	6811827	Smartphone	0	0.568214	1	0.119040	1	0.119040	Devasa	1	1
Kariman Iashiruddin	40	0	karimannashiruddin721@gmail.com	8721553149	Jl. Desa Karang Anyar No. 105, Kec. Megang Sali...	1873087	Jacket	1	0.116752	1	0.024459	0	0.024459	Devasa	1	0

Gambar 4.10 Hasil Labelling

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Berikut *pseudocode labelling* sebagai berikut:



```
# Define threshold for labeling
threshold_jumlah_pembelian = 4 # Customers with 4
or more purchases
threshold_total_harga = dataset['Total
Harga'].quantile(0.75) # Top 25% of spending
threshold_total_pembayaran = dataset['Total
Pembayaran Rata-rata'].quantile(0.75) # Top 25% in
average payment
# Create loyalty label
dataset['Loyalitas'] = (
    (dataset['Jumlah Pembelian'] >=
threshold_jumlah_pembelian) |
    (dataset['Total Harga'] >=
threshold_total_harga) |
    (dataset['Total Pembayaran Rata-rata'] >=
threshold_total_pembayaran)
).astype(int)
```

Gambar 4.11 *Pseudocode Labelling*

Adapun hasil komposisi loyalitas dari proses *labeling* sebagai berikut:

```
Loyalty Label Distribution:
Loyalitas
1    1124
0     876
Name: count, dtype: int64
```

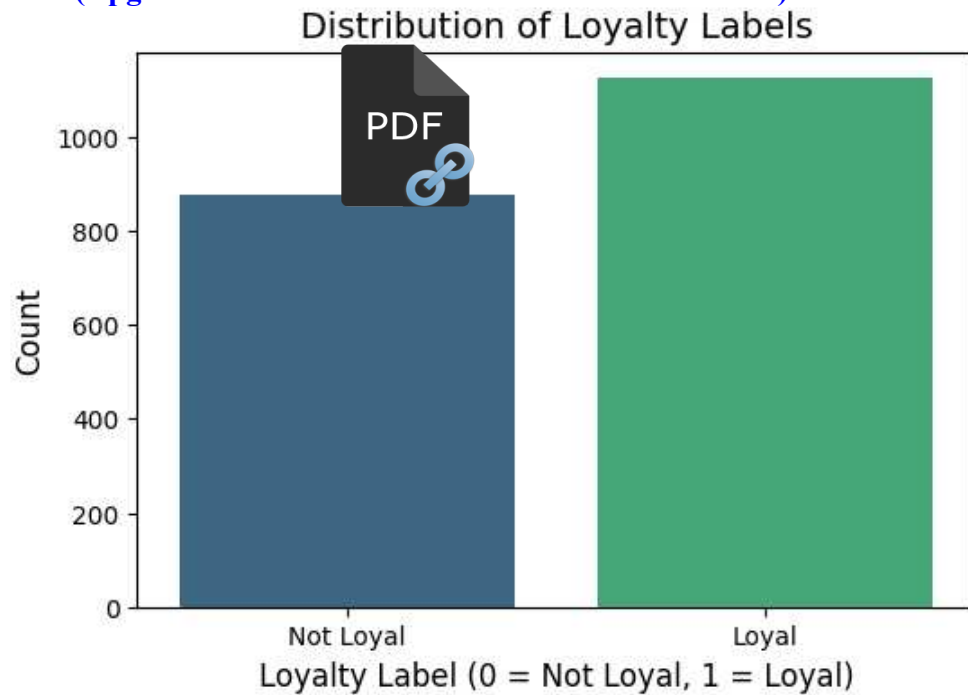
Gambar 4.12 Hasil Komposisi Loyalitas

4.2.4. Visualisasi

Visualisasi ini dirancang untuk menganalisis sebaran pelanggan berdasarkan kategori loyalitas yang telah ditetapkan. Sebaran tersebut disajikan dalam bentuk diagram batang atau grafik lingkaran, yang secara intuitif menggambarkan proporsi antara pelanggan yang loyal dan yang tidak loyal. Penyajian ini mempermudah dalam mengidentifikasi pola-pola utama dalam segmentasi pelanggan.

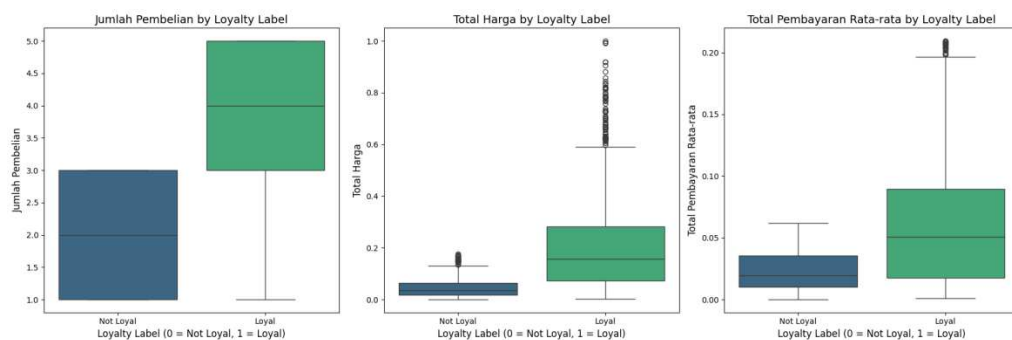
Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



Gambar 4.13 Diagram Loyalitas

Visualisasi berikutnya yang digunakan adalah Boxplot, yang berfungsi untuk menggambarkan distribusi variabel utama seperti jumlah pembelian, total harga, dan rata-rata pembayaran berdasarkan kategori loyalitas. Alat visualisasi ini sangat berguna untuk memahami perbedaan yang signifikan antara pelanggan yang loyal dan yang tidak loyal, serta untuk mengidentifikasi outlier yang mungkin berdampak pada analisis.

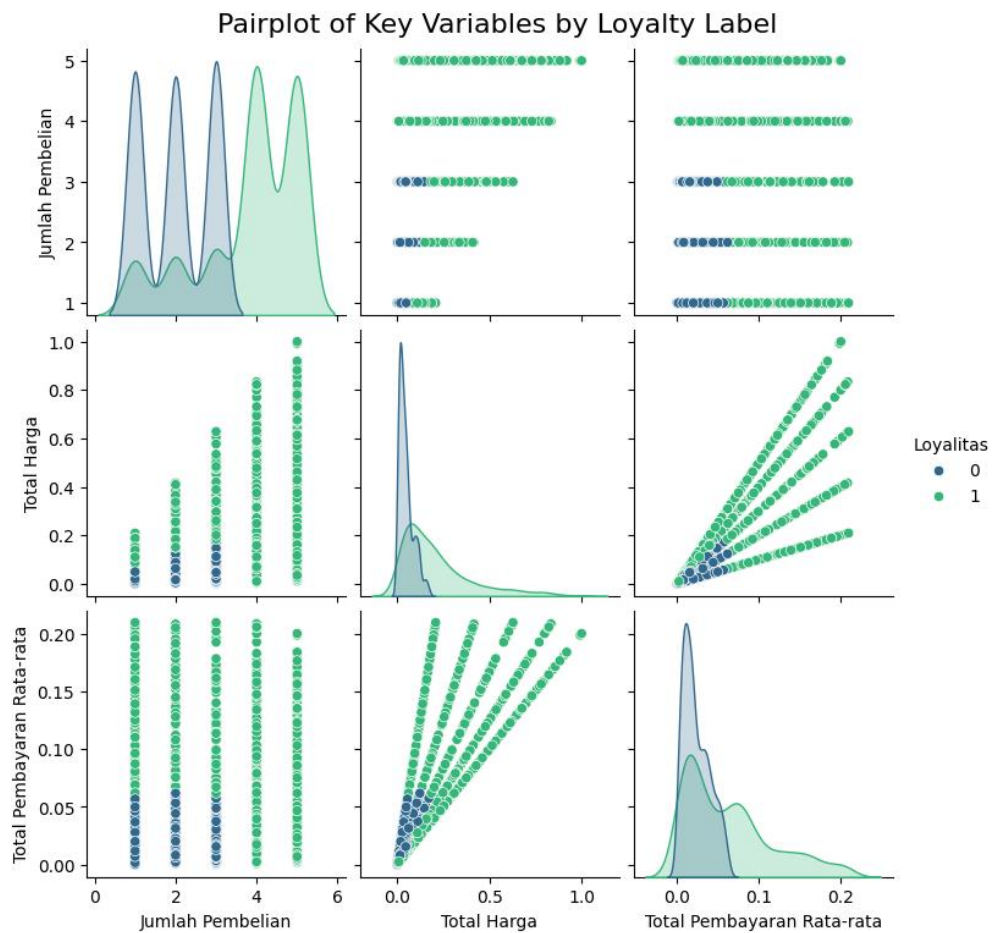


Gambar 4.14 Grafik *Boxplot*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Visualisasi terakhir yang digunakan adalah Pairplot, yang berfungsi untuk menyimpulkan hubungan antara variabel utama dalam dataset, seperti jumlah pembelian, total harga, rata-rata pembayaran, dan label loyalitas. Dengan visualisasi ini, kita dapat mengidentifikasi pola korelasi, baik yang bersifat linier maupun non-linier, serta potensi klustering antarvariabel, yang dapat memberikan pemahaman yang lebih mendalam mengenai karakteristik pelanggan.



Gambar 4.15 Grafik *Parplot*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4.2.5. Klasifikasi

Dalam studi ini, klasifikasi loyalitas pelanggan dilakukan dengan memanfaatkan algoritma *Gradient Boosting*, di mana variabel jumlah pembelian dan total pembelian berfungsi sebagai prediktor. Dataset dipecah menjadi dua bagian, yaitu data pelatihan (80%) dan data pengujian (20%). Untuk mengevaluasi kinerja model, digunakan *Confusion Matrix* yang berguna untuk menganalisis distribusi prediksi, serta *Classification Report* yang digunakan untuk menghitung *accuracy*, *precision*, *recall*, dan *F1-Score*. Proses ini bertujuan untuk memastikan bahwa model dapat memprediksi loyalitas pelanggan dengan efektif dan memberikan hasil yang terukur guna mendukung analisis strategis.

Mulai

```
# Load dataset
file_path = '/KAMPUS/SEMESTER
7/DATA/new/labeled2_shopee2024.csv'
data = pd.read_csv(file_path)
# Encoding variabel kategorikal
label_encoders = {}
for col in ['Nama Customer', 'Jenis Kelamin', 'Email',
'Alamat', 'Kelompok Usia', 'Kelompok Pelanggan', 'Nama
Produk', 'Kategori', 'Metode Pembayaran']:
    le = LabelEncoder()
    data[col] = le.fit_transform(data[col])
    label_encoders[col] = le
# Memilih fitur independen (X) dan target (y)
X = data[['Usia', 'Jenis Kelamin', 'Kelompok Usia',
'Kelompok Pelanggan', 'Kategori', 'Harga', 'Jumlah
Pembelian', 'Total Harga', 'Metode Pembayaran', 'Total
Pembayaran Rata-rata']]
y = data['Loyalitas'] # Misalkan loyalitas adalah
kolom target
# Membagi data menjadi data training dan testing
X_train, X_test, y_train, y_test = train_test_split(X,
y, test_size=0.2, random_state=42)
# Standarisasi fitur numerik
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
# Membuat dan melatih model Gradient Boosting
model = GradientBoostingClassifier(random_state=42)
model.fit(X_train, y_train)
# Prediksi pada data testing
y_pred = model.predict(X_test)
# Evaluasi model
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred,
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```

average='weighted')
recall = recall_score(y_test, y_pred,
average='weight
f1 = f1_score(y_test, y_pred, average='weighted')
# Laporan klasifikasi
print("\n--- Classification Report ---")
print(classification_report(y_test, y_pred))
# Menampilkan confusion matrix secara numerik
print("\n--- Confusion Matrix Numerik ---")
print(confusion_matrix(y_test, y_pred))

# Matriks kebingungan
conf_matrix = confusion_matrix(y_test, y_pred)
sns.heatmap(conf_matrix, annot=True, fmt='d',
cmap='Blues')
# Visualisasi Pentingnya Fitur
feature_importances = model.feature_importances_
features = X.columns
# Tahap Interpretasi Hasil
# 1. Analisis Korelasi
numeric_columns =
data.select_dtypes(include=[np.number])
correlation_matrix = numeric_columns.corr()
# 2. Partial Dependence Plots
# Preprocessing: Features and target
X = data.drop(columns=['Loyalitas'])
y = data['Loyalitas']
X = pd.get_dummies(X, drop_first=True)
# Partial Dependence Plots for important features
features_to_plot = ['Jumlah Pembelian', 'Total Harga',
'Total Pembayaran Rata-rata']
fig, ax = plt.subplots(figsize=(12, 8))
PartialDependenceDisplay.from_estimator(model, X_train,
features=features_to_plot, ax=ax)
# Tahap Visualisasi Data
# 1. Grafik Distribusi Loyalitas Pelanggan
loyalty_distribution = data['Loyalitas'].value_counts()
# 2. Diagram Batang Kategori Produk
product_category_loyalty = data.groupby(['Kategori',
'Loyalitas']).size().unstack()
product_category_loyalty.plot(kind='bar', figsize=(10,
6), stacked=True, color=['orange', 'skyblue'],
alpha=0.8)
# 3. Scatter Plot untuk Hubungan Variabel Numerik
plt.figure(figsize=(10, 6))
sns.scatterplot(data=data, x='Jumlah Pembelian',
y='Total Harga', hue='Loyalitas', alpha=0.8,
palette='cool')
plt.title('Hubungan Jumlah Pembelian dan Total Harga
Berdasarkan Loyalitas', fontsize=14)
plt.xlabel('Jumlah Pembelian', fontsize=12)
plt.ylabel('Total Harga', fontsize=12)
plt.legend(title='Loyalitas')
plt.grid(axis='both', linestyle='--', alpha=0.7)
plt.show()
# 4. Boxplot untuk Distribusi Total Pembayaran Rata-

```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```

rata
plt.figure(figsize=(8, 5))
sns.boxplot(data=data, x='Loyalitas', y='Total
Pembayaran Rata-rata', palette='Set2')
plt.title('Distribusi Total Pembayaran Rata-rata
Berdasarkan Loyalitas', fontsize=14)
plt.xlabel('Loyalitas (0 = Tidak Loyal, 1 = Loyal)',
fontsize=12)
plt.ylabel('Total Pembayaran Rata-rata', fontsize=12)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
# 5. Pie Chart untuk Distribusi Kelompok Usia
age_group_distribution = data['Kelompok
Usia'].value_counts()
age_group_distribution.plot(kind='pie',
autopct='%1.1f%%', figsize=(8, 8), startangle=90,
colors=sns.color_palette('pastel'))
plt.title('Distribusi Kelompok Usia', fontsize=14)
plt.ylabel('')
plt.show()
# 6. Bar Chart untuk Distribusi Kelompok Pelanggan
customer_group_distribution = data.groupby(['Kelompok
Pelanggan', 'Loyalitas']).size().unstack()
customer_group_distribution.plot(kind='bar',
figsize=(10, 6), stacked=True, color=['orange',
'skyblue'], alpha=0.8)
plt.title('Distribusi Loyalitas Berdasarkan Kelompok
Pelanggan', fontsize=14)
plt.xlabel('Kelompok Pelanggan', fontsize=12)
plt.ylabel('Jumlah Pelanggan', fontsize=12)
plt.xticks(rotation=0)
plt.legend(title='Loyalitas', labels=['Tidak Loyal',
'Loyal'])
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()

```

SELESAI

Gambar 4.16 Pseudocode gradient boosting

4.2.6. Evaluasi

Evaluasi hasil klasifikasi mencakup *Confusion Matrix*, *classification Report*, diagram *feature importance*, diagram *heatmap*, diagram loyalitas, diagram *scatter plot*, diagram *boxplot* dan grafik *partial dependence*, memberikan pemahaman mendalam terhadap kinerja model yang di implementasikan menggunakan *Gradient Boosting*, hasil dari evaluasi algoritma *Gradient Boosting* dapat dilihat sebagai berikut:

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```

--- Classification Report ---
              precision    recall  f1-score   support

     0         1.00         1.00         1.00         176
     1         1.00         1.00         1.00         224

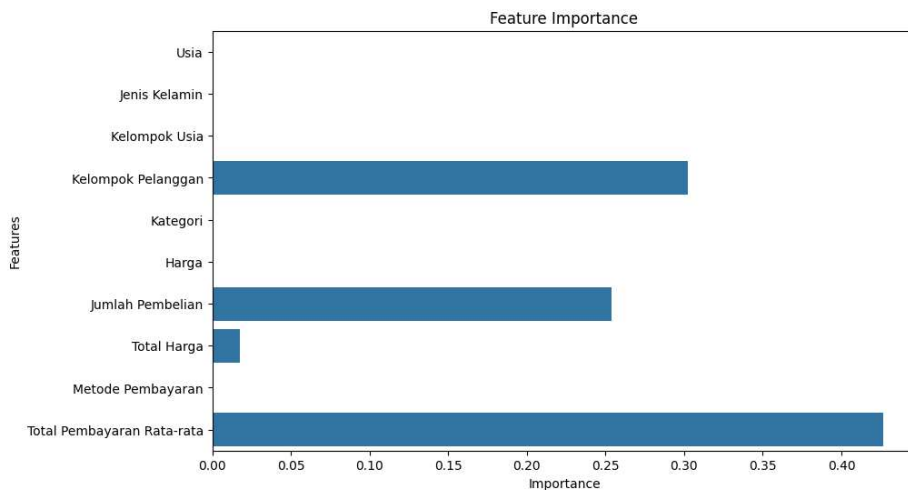
 accuracy          1.00
 macro avg          1.00         1.00         1.00         400
 weighted avg       1.00         1.00         1.00         400
  
```

Gambar 4.17 Hasil Evaluasi *Classification Report*

```

--- Confusion Matrix Numerik ---
[[176  0]
 [ 0 224]]
  
```

Gambar 4.18 Hasil *Confusion Matrix*



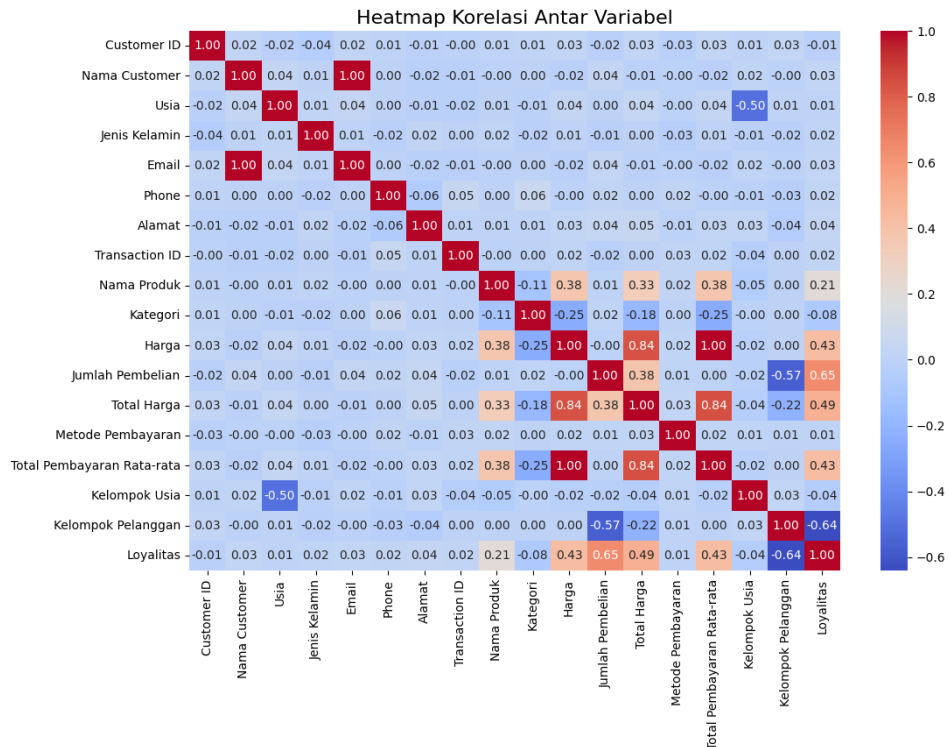
Gambar 4.19 Diagram *Feature Importance*

Diagram ini menunjukkan tingkat relevansi setiap fitur dalam model pembelajaran mesin, di mana fitur Rata-rata Total Pembayaran memiliki bobot tertinggi, mengindikasikan dampak besar dalam

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

mempengaruhi prediksi model. Fitur lain yang juga cukup berpengaruh adalah Kelompok Pelanggan dan Jumlah Pembelian, sementara fitur seperti Jenis Kelamin, Usia, dan Alamat Kategori tidak memiliki signifikansi yang berarti dalam konteks prediksi model ini.



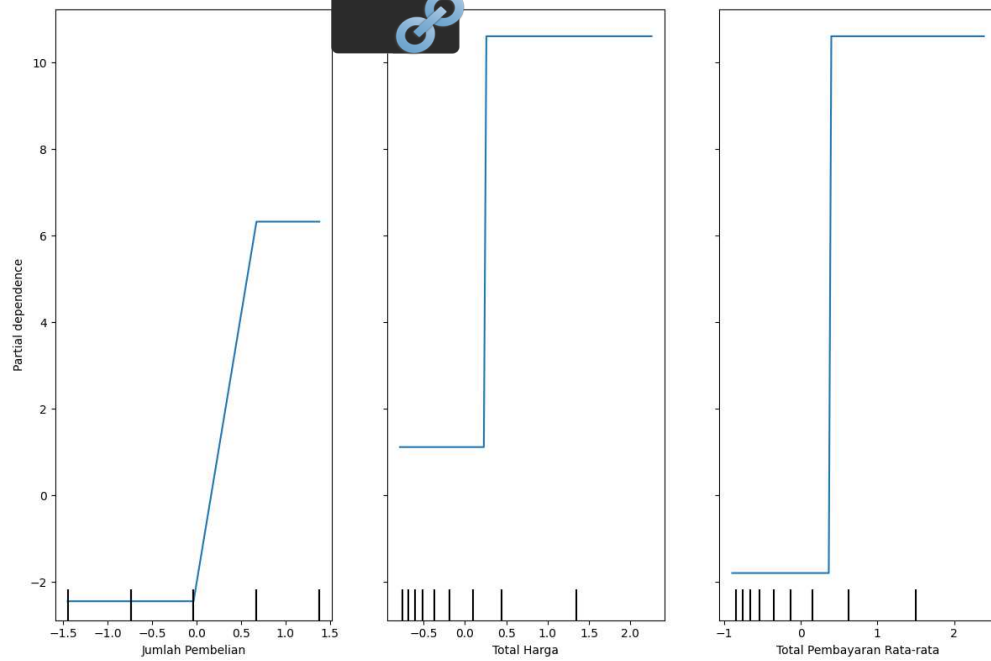
Gambar 4.20 Diagram *Heatmap*

Diagram ini menggambarkan hubungan antar variabel dalam kumpulan data, dengan nilai korelasi berkisar antara -1 hingga 1. Korelasi positif yang tinggi (dekat dengan 1) menunjukkan adanya hubungan searah antara dua variabel, sedangkan korelasi negatif yang tinggi (dekat dengan -1) menunjukkan adanya hubungan yang berlawanan. Nilai yang mendekati 0 menandakan adanya hubungan yang lemah atau bahkan tidak ada hubungan sama sekali. Variabel seperti Jumlah Pembelian, Total Harga, dan Rata-rata Total Pembayaran menunjukkan korelasi yang tinggi (dekat dengan 1), yang mengindikasikan adanya hubungan yang kuat di antara fitur-fitur transaksi tersebut. Di sisi lain, Kelompok Usia menunjukkan

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

korelasi negatif yang moderat (-0.50) dengan Usia, yang menunjukkan adanya hubungan yang cekat tetapi dengan arah yang berlawanan.

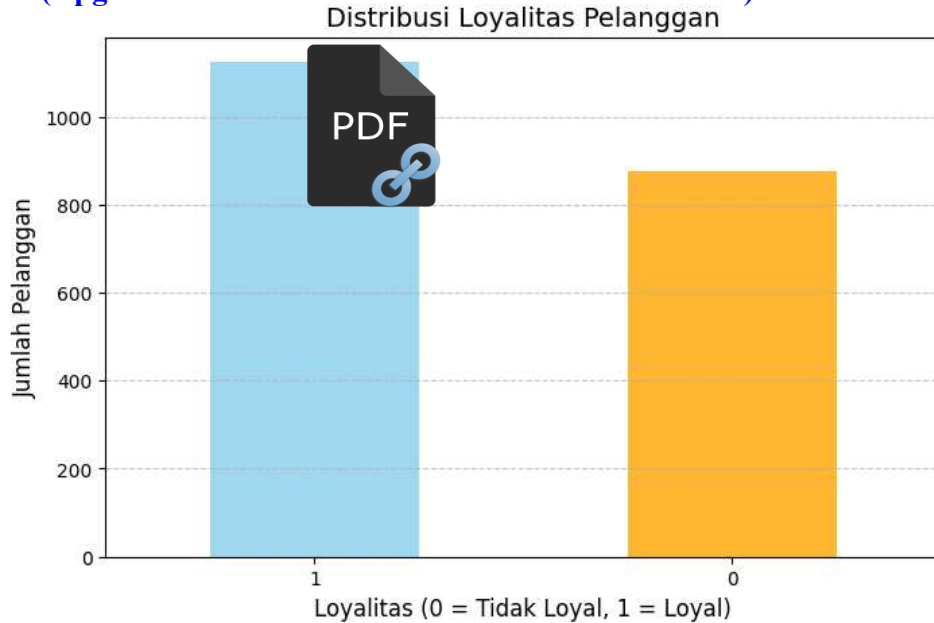


Gambar 4.21 Grafik Partial Dependence Plot

Grafik ini menampilkan Partial Dependence Plots (PDP) untuk tiga fitur utama: Jumlah Pembelian, Total Harga, dan Rata-rata Total Pembayaran. Fitur Jumlah Pembelian menunjukkan hubungan linier, di mana dampaknya meningkat secara signifikan setelah mencapai nilai nol. Untuk fitur Total Harga, dampaknya meningkat secara drastis di awal, namun kemudian cenderung stabil. Pola yang serupa juga terlihat pada Rata-rata Total Pembayaran, di mana dampak yang kuat terlihat pada awalnya dan kemudian menjadi konstan.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

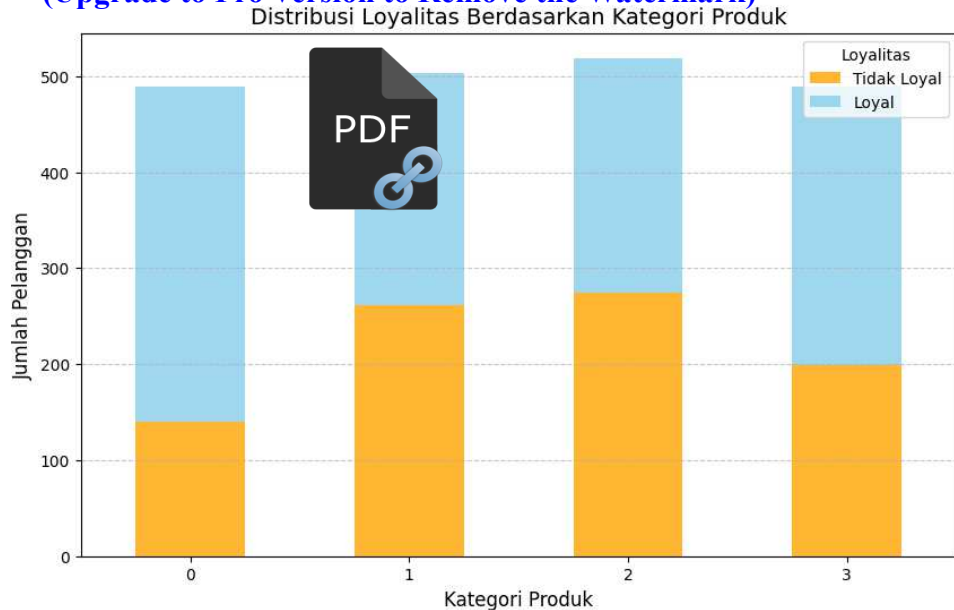


Gambar 4.22 Diagram Loyalitas Pelanggan

Diagram tersebut menggambarkan sebaran loyalitas pelanggan yang terbagi dalam dua kelompok, yaitu 0 (Tidak Loyal) yang diwakili oleh warna oranye, dan 1 (Loyal) yang diwakili oleh warna biru muda. Dari grafik tersebut, dapat dilihat bahwa jumlah pelanggan yang loyal (kategori 1) lebih banyak dibandingkan dengan pelanggan yang tidak loyal (kategori 0), hal ini terlihat dari ketinggian batang berwarna biru yang melebihi batang berwarna oranye.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

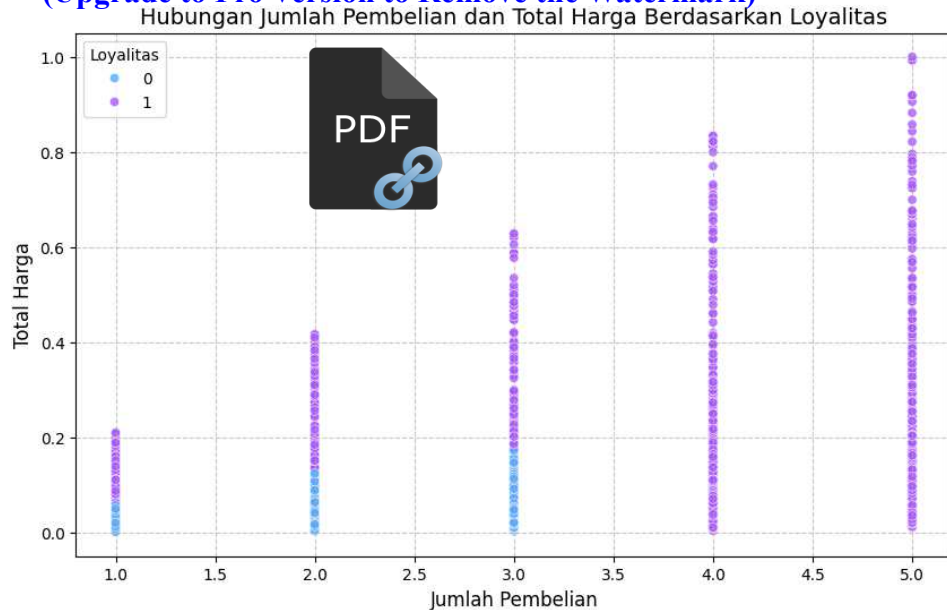


Gambar 4.23 Diagram Loyalitas Berdasarkan Kategori Produk

Diagram di atas menggambarkan penyebaran loyalitas pelanggan menurut kategori produk, yaitu kategori 0 (elektronik), 1 (fashion), 2 (kecantikan), dan 3 (peralatan rumah tangga). Pada diagram tersebut dibagi menjadi dua bagian: warna oranye melambangkan pelanggan yang tidak loyal, sementara warna biru muda melambangkan pelanggan yang loyal. Dari grafik ini, dapat dilihat bahwa di setiap kategori produk, jumlah pelanggan yang loyal selalu lebih banyak dibandingkan dengan pelanggan yang tidak loyal. Penyebaran loyalitas cukup merata di seluruh kategori produk, dengan jumlah pelanggan loyal yang cenderung stabil di setiap kategori. Di sisi lain, proporsi pelanggan yang tidak loyal sedikit bervariasi antar kategori, tetapi secara keseluruhan tetap lebih rendah dibandingkan dengan jumlah pelanggan yang loyal.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

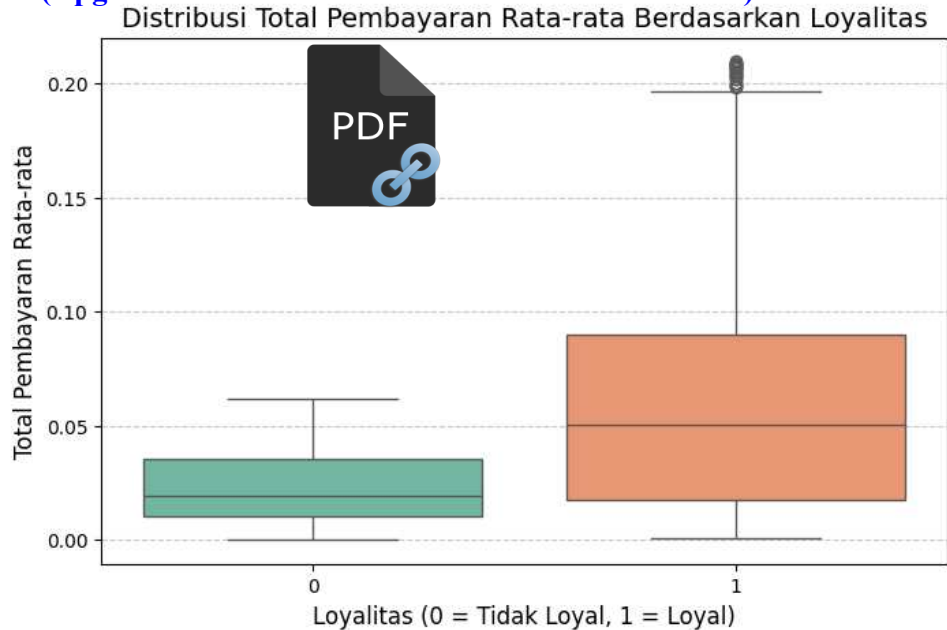


Gambar 4.24 Grafik *Scatter Plot*

Grafik *scatter plot* di atas menggambarkan hubungan antara jumlah pembelian dan total harga, dengan loyalitas pelanggan sebagai penanda warna: titik biru melambangkan pelanggan yang tidak loyal (loyalitas 0), sedangkan titik ungu melambangkan pelanggan yang loyal (loyalitas 1). Dari grafik tersebut, terlihat bahwa pelanggan loyal (ungu) lebih banyak muncul di seluruh rentang jumlah pembelian dan total harga. Selain itu, semakin tinggi jumlah pembelian atau total harga, mayoritas pelanggan cenderung menunjukkan loyalitas. Sebaliknya, pelanggan yang tidak loyal (biru) lebih sering ditemukan pada jumlah pembelian dan total harga yang cenderung lebih rendah.

Protected by PDF Anti-Copy Free

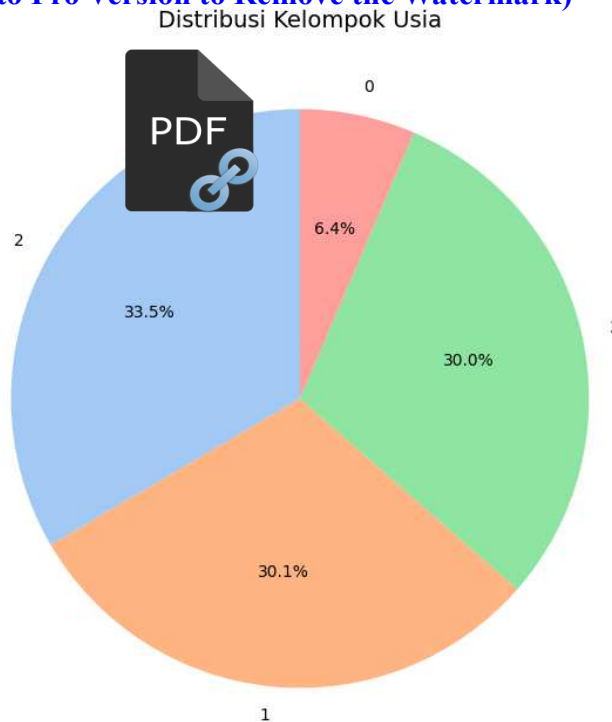
(Upgrade to Pro Version to Remove the Watermark)



Gambar 4.25 Boxplot perbandingan

Analisis distribusi nilai menunjukkan bahwa kelompok pelanggan yang tidak loyal (0) memiliki rata-rata total pembayaran yang lebih rendah jika dibandingkan dengan kelompok pelanggan yang loyal (1). Rentang rata-rata pembayaran pada pelanggan loyal lebih bervariasi dan terdapat beberapa nilai ekstrim (outliers), yang menunjukkan adanya pelanggan dengan pembayaran yang jauh lebih tinggi dari rata-rata. Dari perspektif pusat data, median pembayaran rata-rata untuk pelanggan loyal lebih tinggi dibandingkan dengan pelanggan yang tidak loyal, yang menunjukkan bahwa pelanggan loyal cenderung memberikan kontribusi yang lebih signifikan terhadap total pembayaran. Selain itu, penyebaran data dalam kelompok pelanggan loyal lebih luas dibandingkan dengan kelompok tidak loyal, yang dapat dilihat dari ukuran box dan whisker yang lebih panjang.

Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)

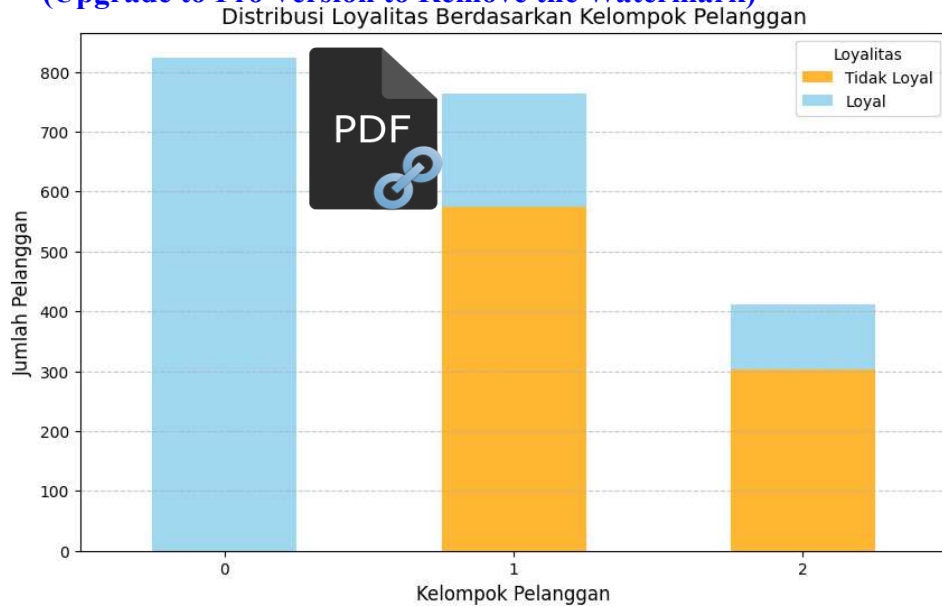


Gambar 4.26 Diagram Distribusi Kelompok Usia

Distribusi usia menunjukkan bahwa kelompok 2, yang terdiri dari dewasa muda berusia 25-35 tahun, memiliki persentase tertinggi, yaitu 33,5%. Sementara itu, kelompok 1 atau remaja berusia 17-25 tahun dan kelompok 3, yang mencakup dewasa berusia 35-50 tahun, memiliki persentase yang hampir setara, masing-masing sebesar 30,1% dan 30,0%, yang sedikit lebih rendah daripada kelompok usia 2. Di sisi lain, kelompok 0, yang terdiri dari anak-anak berusia 0-16 tahun, mencatatkan persentase terendah, yaitu 6,4%, menunjukkan bahwa kelompok ini tergolong sebagai minoritas dalam keseluruhan data.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



Gambar 4.27 Diagram Loyalitas Berdasarkan Kelompok Pelanggan

Kelompok pelanggan 0 didominasi oleh pelanggan yang loyal, dengan jumlah yang jauh lebih banyak dibandingkan dengan pelanggan yang tidak loyal. Sementara itu, kelompok pelanggan 1 menunjukkan distribusi yang lebih seimbang antara pelanggan loyal dan yang tidak loyal, meskipun jumlah pelanggan tidak loyal sedikit lebih banyak. Berbeda dengan itu, kelompok pelanggan 2 memiliki proporsi yang lebih tinggi dari pelanggan tidak loyal dibandingkan dengan pelanggan loyal. Dari segi karakteristik, kelompok 0 terdiri dari pelanggan dengan frekuensi pembelian rendah dan total nilai yang kecil (pembeli sesekali dengan nilai rendah), kelompok 1 mencakup pelanggan dengan frekuensi dan nilai pembelian yang sedang, sedangkan kelompok 2 terdiri dari pelanggan dengan frekuensi pembelian tinggi dan nilai transaksi yang besar (pelanggan loyal dengan belanja yang signifikan).

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4.3. Pembahasan

Masalah yang diangkat dalam penelitian ini adalah analisis klasifikasi tingkat loyalitas pelanggan platform e-commerce Shopee dengan menggunakan algoritma *Gradient Boosting*. Tujuannya adalah untuk mengidentifikasi tingkat loyalitas pelanggan berdasarkan data yang telah dikumpulkan. Tahap awal dari penelitian ini mencakup pengumpulan data melalui API Shopee, dengan penekanan pada informasi transaksi dan perilaku pengguna, untuk memperoleh pemahaman yang lebih mendalam mengenai faktor-faktor yang memengaruhi loyalitas pelanggan. Data yang berhasil dikumpulkan akan berkontribusi dalam pengembangan model klasifikasi yang tepat untuk membedakan antara pelanggan yang loyal dan yang tidak. Setelah pengumpulan data selesai, langkah berikutnya adalah tahap pra-pemrosesan yang menghasilkan data yang telah dibersihkan. Setelah dilakukan *preprocessing* tahapan berikutnya adalah *labeling* adapun hasil komposisi dari *labeling* dapat dilihat pada tabel

Tabel 4.1 Label Loyalitas

Atribut	Loyal	Tidak Loyal
Loyalitas	1124	876

Tahap terakhir adalah penerapan klasifikasi dengan menggunakan algoritma *Gradient Boosting*, di mana teknik evaluasi yang digunakan adalah Confusion Matrix dan laporan klasifikasi. Hasil evaluasi dapat dilihat di bawah ini:

Tabel 4.2 Klasifikasi *Gradient Boosting*


Loyalitas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Accuracy</i>
Tidak Loyal	100%	100%	100%	100%
Loyal	100%	100%	100%	

Protected by PDF Anti-Copy Free

[\(Upgrade to Pro Version to Remove the Watermark\)](#)

Berdasarkan tabel evaluasi model yang menyajikan metrik seperti Precision, Recall, F1-Score, dan Accuracy, hasil yang diperoleh untuk semua kategori (Loyal dan Tidak Loyal) mencapai 100%. Hal ini menandakan bahwa model klasifikasi yang diterapkan memiliki kinerja yang sangat baik dalam mengenali kategori tersebut tanpa kesalahan. Semua prediksi yang dihasilkan oleh model, baik yang positif maupun negatif, sesuai dengan nilai yang sebenarnya.

BAB V
KESIMPULAN DAN SARAN



5.1. Kesimpulan

Penelitian ini telah berhasil mencapai tujuan utamanya, yaitu mengidentifikasi sejumlah faktor kunci yang berpengaruh terhadap loyalitas pelanggan Shopee yang berada di wilayah Musi Rawas. Selain itu, penelitian ini juga mengembangkan sebuah model klasifikasi yang bertujuan untuk mengkategorikan tingkat loyalitas pelanggan dengan memanfaatkan algoritma Gradient Boosting. Model yang dihasilkan dari penelitian ini menunjukkan performa yang sangat memuaskan, dengan tingkat akurasi mencapai 100% dalam mengklasifikasikan pelanggan ke dalam kategori loyal dan tidak loyal.

Dari hasil analisis yang dilakukan, terungkap bahwa terdapat beberapa faktor yang memiliki pengaruh paling signifikan dalam proses prediksi loyalitas pelanggan. Di antara faktor-faktor tersebut, Total Pembayaran Rata-rata, Kelompok Pelanggan, dan Jumlah Pembelian menjadi faktor-faktor yang paling dominan. Selain itu, analisis juga menunjukkan adanya korelasi yang signifikan antar berbagai variabel yang diteliti. Misalnya, ditemukan adanya korelasi positif antara variabel Harga, Total Harga, dan Total Pembayaran Rata-rata, yang menunjukkan bahwa peningkatan pada variabel-variabel tersebut berhubungan langsung dengan loyalitas pelanggan. Sebaliknya, terdapat juga korelasi negatif antara Kelompok Usia dan Usia, yang mengindikasikan bahwa perubahan dalam kelompok usia dapat mempengaruhi tingkat loyalitas pelanggan.

Hasil penelitian ini memberikan pemahaman mendalam tentang perilaku pelanggan di Shopee, serta bagaimana faktor-faktor tertentu mempengaruhi keputusan pembelian dan loyalitas mereka. Dengan pemahaman ini, perusahaan dapat menerapkan model klasifikasi yang telah dikembangkan untuk meningkatkan strategi pemasaran. Ini akan mendukung Shopee dalam

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

memperkuat hubungan dengan pelanggan dan meningkatkan kepuasan serta loyalitas mereka di masa depan



5.2. Saran


Untuk meningkatkan hasil penelitian pada tahap selanjutnya, terdapat beberapa rekomendasi yang dapat diberikan kepada peneliti berikutnya. Rekomendasi ini diharapkan dapat mendukung penelitian dalam mencapai hasil yang lebih baik, yaitu:

- a. Memanfaatkan dataset yang lebih besar dan variatif secara geografis untuk meningkatkan kemampuan generalisasi hasil penelitian serta representasi populasi pelanggan
- b. Mencoba algoritma lain, seperti Random Forest atau Neural Network, untuk memberikan wawasan tambahan atau meningkatkan kinerja model klasifikasi yang dihasilkan
- c. Memperluas analisis ke aspek-aspek lain, seperti tren topik pelanggan, frekuensi interaksi dengan aplikasi, atau hubungan antara sentimen pelanggan dan faktor eksternal tertentu, untuk memberikan pemahaman yang lebih komprehensif mengenai loyalitas pelanggan.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR PUSTAKA

- 
- [1] P. L. Isnaini, P. Studi, E. U. Nurul, and J. Probolinggo, “Pengaruh Penggunaan Quick Response Code Indonesian Standard (QRIS) Oleh UMKM Terhadap Pendapatan,” vol. 4, no. 1, pp. 1–11, 2023.
- [2] P. T. Simangungsong, “Dampak Atribut Belanja Dan Pengalaman Belanja Online Terhadap Loyalitas Pelanggan Dengan Kepuasan Sebagai Mediasi Di E-Commerce Shoppe,” pp. 1–42, 2023.
- [3] A. Felix and G. D. Rembulan, “Digital Transformation and the Customer Experience: Enhancing Engagement and Loyalty,” *Entrep. J. Bisnis Manaj. dan Kewirausahaan*, vol. 4, no. 03, pp. 228–240, 2023, doi: 10.31949/entrepreneur.v4i03.6195.
- [4] A. Haksanggulawan, I. Hajar, and A. Putera, “Analisis Mekanisme Pasar Dalam Pasar Tradisional Di Indonesia,” *J. Ekon. Manaj. dan Akunt. Sekol. Tinggi Ilmu Ekon. Enam-Enam Kendari*, vol. 1, no. 2, pp. 401–407, 2023, [Online]. Available: <https://doi.org/10.572349/neraca.v1i2.163%0Ahttps://jurnal.kolibi.org/index.php/neraca/article/view/163>
- [5] P. P. Nugraha, “Korelasi Ramalan Joyoboyo Pasar Ilang Kumandange Terhadap Pemasaran Digital Di Era Marketing 4.0 Di Indonesia,” *Bangun Rekaprima*, vol. 9, no. 1, p. 75, 2023, doi: 10.32497/bangunrekaprima.v9i1.4475.
- [6] D. Anugrah Pratama, I. Rizal Mutaqin, and K. Rafael Manuela, “Analisis Terjadinya Kanker Paru-Paru Pada Pasien Menggunakan Decision Tree: Penerapan Algoritma C4.5 Dan RapidMiner Untuk Menentukan Risiko Kanker Pada Gejala Pasien,” *Jtmei*, vol. 2, no. 4, pp. 156–170, 2023, [Online]. Available: <https://doi.org/10.55606/jtmei.v2i4.3004>
- [7] Yuyut Prayuti, “Dinamika Perlindungan Hukum Konsumen di Era Digital: Analisis Hukum Terhadap Praktik E-Commerce dan Perlindungan Data Konsumen di Indonesia,” *J. Interpret. Huk.*, vol. 5, no. 1, pp. 903–913, 2024, doi: 10.22225/juinhum.5.1.8482.903-913.
- [8] F. R. Sari, “Pengaruh Biaya Promosi Terhadap Volume Penjualan di Toko Pusat Oleh - Oleh Pusako Minang Tabek Patah,” *Skripsi*, 2020.
- [9] M. S. Ummah, “No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title,” *Sustain.*, vol. 11, no. 1, pp. 1–14, 2019, [Online]. Available: <http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/3053204>

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

84_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI

- [10] T. K. E-commerce, A. P. Wulandari, and A. A. Azzahra, "10.8734/Kohesi.v1i2.36 PDF", no. 1, pp. 1–14, 2024.
- [11] A. Sobandi, "Pengolahan dalam Sistem Informasi," *Manajerial J. Manaj. dan Sist. Inf.*, vol. 1, no. 1, pp. 89–95, 2002, [Online]. Available: <https://ejournal.upi.edu/index.php/manajerial/article/view/16450>
- [12] I Putu Ramayasa, I Gusti Ayu Desi Saryanti, I Komang Dharmendra, and Edwar, "Perbandingan Metode Vektorisasi Pada Analisa Sentiment, Studi Kasus : Cyberbullying Pada Komentar Instagram," *J. Teknol. Inf. dan Komput.*, vol. 9, no. 5, pp. 505–512, 2023, doi: 10.36002/jutik.v9i5.2645.
- [13] Sudarta, "濟無No Title No Title No Title," vol. 16, no. 1, pp. 1–23, 2022.
- [14] K. JASMINE, "濟無No Title No Title No Title," *Penambahan Natrium Benzoat Dan Kalium Sorbat Dan Kecepatan Pengadukan Sebagai Upaya Penghambatan Reaksi Inversi Pada Nira Tebu*, 2014.
- [15] M. A. Rahman, A. Warokka, and R. Y. Priyati, "Minat Penggunaan Cashless Payment Sistem Dompot Digital Pada Mahasiswa Di Papua Barat," *Sci. J. Reflect. Econ. Accounting, Manag. Bus.*, vol. 7, no. 3, pp. 852–867, 2024, doi: 10.37481/sjr.v7i3.909.
- [16] F. Ramdani and I. Q. Utami, *Pengantar Data Science*. 2022.
- [17] R. Adolph, 濟無No Title No Title No Title. 2016.
- [18] W. Sulistiyowati, "Buku Ajar Statistika Dasar," *Buku Ajar Stat. Dasar*, vol. 14, no. 1, pp. 15–31, 2017, doi: 10.21070/2017/978-979-3401-73-7.
- [19] F. S. Nadiansyah, R. Indrawan, and S. Almujab, "Analisis Strategi Promosi E-Commerce Shopee Yang," *Oikos*, vol. 6, no. 1, pp. 62–71, 2022.
- [20] I. N. Science, "INTERNATIONAL JOURNAL OF Prediction of Customer Loyalty in E-Commerce using Machine Learning Models," vol. 7, no. 9, 2024, doi: 10.15680/IJMRSET.2024.0709018.

Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)



LAMPIRAN

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Lampiran 1 Pengesahan Judul

UNIVERSITAS BINA INSAN
PENDIDIKAN DWI TUNGGAL PALEMBANG

Formulir Pengajuan Judul Skripsi
Program Studi Sistem Informasi

Nama : Feri Basofi
NIM : 2102030046
Alamat : Desa Kalibening Dusun IV Kec.Tugumulyo Kab. Musi Rawas
No.Hp : 0857-8873-2413

Rumusan Masalah 1 : Bagaimana penggunaan live hosting MySQL dalam sistem informasi penilaian siswa dapat meningkatkan efisiensi dan akurasi penyimpanan data?
Judul 1 ~~A~~ : Sistem Informasi Penilaian Siswa Berbasis Desktop dengan Live Hosting MySQL di SD IT Al-Qudwah Musi Rawas


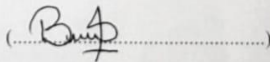
Rumusan Masalah 2 : Bagaimana menerapkan algoritma Gradient Boosting untuk mengklasifikasikan loyalitas pelanggan di platform e-commerce Shopee?
Judul 2 ~~...~~ ACC : Penerapan Data Mining untuk Klasifikasi Loyalitas Pelanggan E-Commerce Shopee Menggunakan Algoritma Gradient Boosting

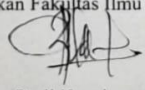
Rumusan Masalah 3 : Bagaimana penerapan algoritma XGBoost dalam mengklasifikasikan siswa berprestasi di bidang kepramukaan?
Judul 3 : Implementasi Algoritma XGBoost untuk Klasifikasi Siswa Berprestasi di Bidang Kepramukaan pada Kwartir Cabang Musi Rawas Keluarga

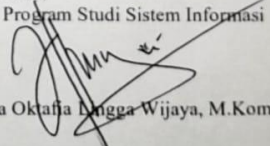
Diusulkan Judul Nomor (satu) / 2 (Dua) / 3 (Tiga)*

Lubuklinggau, September 2024
Mahasiswa yang mengusulkan,

(Feri Basofi)

Menyetujui Dosen Pembimbing,
Pembimbing 1 (Budi Santoso, M.Kom) 
Pembimbing 2 (Bunga Intan, M.Kom) 

Mengesahkan,
Dekan Fakultas Ilmu Teknik 
(Dr. Rudi Kurniawan, ST., M.Kom)


Mengetahui,
Ketua Program Studi Sistem Informasi 
(Harma Oktavia Lingga Wijaya, M.Kom)

0733-4553932 (Rektorat Universitas) 0812-1826-6228 (Marketing UNIVBI)
0733-3280300 (Bina Insan) 0852-3151-5800 (Admin UNIVBI)
0733-3280200 (Pascasarjana) Admin@univbinainsan.ac.id univbinainsan.ac.id - pasca.univbinainsan.ac.id

Protected by PDF Anti-Copy Free





(Upgrade to Pro Version to Remove the Watermark)

Lampiran 2 Lembar Bimbingan


UNIVERSITAS BINA INSAN
 FAKULTAS ILMU TEKNIK
 Lubuk Rukung Kec. Lubuklinggau Selatan 1 Kota Lubuklinggau Provinsi Sumatera Selatan

LEMBAR BIMBINGAN SKRIPSI

Nama : Feri BAsofi
 Nim : 2102030046
 Program Studi : Sistem Informasi
 Pembimbing 1 : Budi Santoso, M.Kom
 Pembimbing 2 : Bunga Intan, M.Kom
 Judul : Penerapan Data Mining untuk Klasifikasi Loyalitas Pelanggan E-Commerce Shopee menggunakan algoritma gradient boosting


NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
1	11/01/2025	skripsi	<ul style="list-style-type: none"> Perbaiki format spasi pada daftar isi lengkapi daftar tabel, daftar gambar dan daftar lampiran Perbaiki format nama halaman setiap awal bab nomor tabel tidak di bold 		
2	13/01/2025	skripsi	<ul style="list-style-type: none"> Perbaiki format daftar pustaka sertakan program pengolahan data lengkapi lampiran. 		
3	15/01/2025	skripsi	<ul style="list-style-type: none"> Perbaiki kalimat pada abstrak max 500 kata termasuk masalah dan hasil lengkapi pada hasil debat & 		
4	16/01/2025	skripsi	AC. 5 halaman layout PI		

Lubuklinggau, 2025
Ketua Program Studi Sistem Infomasi

(Harma Oktafia Lingga W., M.Kom)

0733-4553932 (Rektorat Universitas) 0812-1826-6228 (Marketing UNIVBI)
 0733-3280300 (Bina Insan) 0852-3151-5800 (Admin UNIVBI)
 0733-3280200 (Pascasarjana) Admin@univbinainsan.ac.id univbinainsan.ac.id - pasca.univbinainsan.ac

Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)

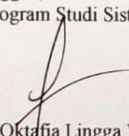


YAYASAN DWI TUNGGAL PALEMBANG
UNIVERSITAS BINA INSAN
 FAKULTAS ILMU TEKNIK
Jalan Jenderal Besar
Lubuk Kuning, Kec. Lubuklinggau Selatan I, Kota Lubuklinggau, Provinsi Sumatera Selatan

LEMBAR Bimbingan SKRIPSI

Nama : Feri BAsofi
 Nim : 2102030046
 Program Studi : Sistem Informasi
 Pembimbing 1 : Budi Santoso, M.Kom
 Pembimbing 2 : Bunga Intan, M.Kom
 Judul : Penerapan Data Mining untuk Klasifikasi Loyalitas Pelanggan E-Commerce Shopee menggunakan algoritma gradient boosting

NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
	16/2025 1		- Perbaiki metode klasifikasi nya sawikan dg data sbli - flowchart pengolahan datanya	[Signature]	
	17/2025 1		- lengkapi Berkas	[Signature]	
	16/2025		- Acc Silakan usun	[Signature]	

Lubuklinggau,2025
 Ketua Program Studi Sistem Infomasi

 (Harna Oktafia Lingga W., M.Kom)

0733-4553932 (Rektorat Universitas) 0812-1826-6228 (Marketing UNIVBI)
 0733-3280300 (Bina Insan) 0852-3151-5800 (Admin UNIVBI)
 0733-3280200 (Pascasarjana) Admin@univbinaisan.ac.id univbinaisan.ac.id - pasca.univbinaisan.ac

Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)

Lampiran 3 Dataset



Customer ID	Nama Customer	Usia	Jenis Kelamin	Email	Alamat
855409	Hafshah Wibowo	19	Perempuan	hafshahwibowo929@gmail.com	Jl. Desa Muara Kelingi No. 1, Kec. Muara Kelingi, Kab. Musi Rawas, Sumatera Selatan
617884	Taufik Sudiati	16	Perempuan	taufiksudiati275@gmail.com	Jl. A. Yani No. 92, Kec. Muara Beliti, Kab. Musi Rawas, Sumatera Selatan
336002	Wadi Permadi	26	Perempuan	wadipermedi321@gmail.com	Jl. Desa Purwodadi No. 48, Kec. Purwodadi, Kab. Musi Rawas, Sumatera Selatan
555351	Bajragin Yuliarti	35	Laki-Laki	bajraginyuliarti195@gmail.com	Jl. Desa Muara Kelingi No. 23, Kec. Muara Kelingi, Kab. Musi Rawas, Sumatera Selatan
742947	Kariman Nashirudin	40	Laki-Laki	karimannashiruddin721@gmail.com	Jl. Desa Karang Anyar No. 105, Kec. Megang Sakti, Kab. Musi Rawas, Sumatera Selatan

Transaction ID	Customer ID	Nama Customer	Nama Produk	Harga	Jumlah Pembelian	Total Harga	Metode Pembayaran
3290956	855409	Hafshah Wibowo	Headphone	1342810	4	5371240	COD
8512438	617884	Taufik Sudiati	T-Shirt	76557	4	306228	COD
8143910	336002	Wadi Permadi	Lipstick	194619	4	778476	Transfer Bank
6811827	555351	Bajragin Yuliarti	Smartphone	2839961	1	2839961	Seabank
1873087	742947	Kariman Nashirudin	Jacket	599700	1	599700	COD