PENERAPAN METODE ALGORITMA EXTREME GRADIENT BOOSTING DALAM MEMPREDIKSI PENJUALAN THRIFTING PADA TOKO CCHASE

SKRIPSI

DISUSUN OLEH

M. ICHWAN GIFARI

NPM. 2109020161



PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

MEDAN

2025

PENERAPAN METODE ALGORITMA EXTREME ALGORITMA EXTREME GRADIENT BOOSTING DALAM MEMPREDIKSI PENJUALAN THRIFTING PADA TOKO CCHASE

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) dalam Program Studi Teknologi Informasi pada Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara

M. ICHWAN GIFARI
NPM. 2109020161

PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025

LEMBAR PENGESAHAN

Judul Skripsi	: PENERAPAN METODE ALGORITMA EXTREME GRADIENT BOOSTING DALAM MEMPREDIKSI PENJUALAN THRIFTING PADA TOKO CCHASE
Nama Mahasiswa	: M. Ichwan Gifari
NPM	: 2109020161
Program Studi	: Teknologi Informasi

Menyetujui
Komisi Pembimbing

(Yoshida Sary S.Kom, S.E., M.Kom)
NIDN. 0105067603

Ketua Program Studi	Dekan
Fatma Sari Hutagalung, S.Kom., M.Kom.	Dr. Al-Khowarizmi, S.Kom., M.Kom.
NIDN. 0117019301	NIDN. 0127099201

PERNYATAAN ORISINALITIAS

Penerapan Metode Algoritma Extreme Gradient Boosting Dalam Memprediksi Penjualan Thrifting Pada Toko Cchase

SKRIPSI

Saya menyatakan bahwa karya tulis ini hasilkarya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 16 september 2025

Yang Membuat Pernyataan



M. Ichwan Gifari

NPM. 2109020161

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademi Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan di bawah ini:

Nama	: M. Ichwan Gifari
NPM	: 2109020161
Program Studi	: Teknologi Informasi
Karya Ilmiah	: Skripsi

Demi oengembangan ilmu pengetahuan, menyetujui untuk membetikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif ini, *(Non-Exclusive Royalty Free Right)* atas penelitian skripsi yang berjudul:

PENERAPAN METODE ALGORITMA EXTREME GRADIENT BOOSTING DALAM MEMPREDIKSI PENJUALAN THRIFTING PADA TOKO CCHASE

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih mediah, memformat, mengelola dalam bentuk database, merawat dan mempulikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau pemilik hak cipta.

Demikian Pernyataan ini dibuat dengan sebenartnya.

Medan, 16 september 2025
Yang Membuat Pernyataan



M. Ichwan Gifari
NPM. 2109020161

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap	: Muhammad Ichwan Gifari
Tempat dan Tanggal Lahir	: MEDAN, 12 FEBRUARI 2004
Alamat Rumah	: JL. Air Bersih Blok 3 No,12
Telepon/Faks/HP	087777440149
E-mail	: ichwangifari07@gmail.com
Instansi Tempat Kerja	: -
Alamat Kantor	: -

DATA PENDIDIKAN

SD	: SD PERGURUAN ERIA	TAMAT: 2015
SMP	: SWASTA AL-ULUM	TAMAT: 2018
SMA	: SWASTA AL-ULUM	TAMAT: 2021

KATA PENGANTAR



Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa, atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "PENERAPAN METODE ALGORITMA EXTREME GRADIENT BOOSTING DALAM MEMPREDIKSI PENJUALAN THRIFTING PADA TOKO CCHASE" ini dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknologi Informasi, Fakultas Ilmu Komputer & Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara.

Penulisan skripsi ini tidak lepas dari bantuan, dukungan, serta bimbingan dari berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU).
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Fatma Sari Hutagalung, S.Kom., M.Kom. Ketua Program Studi Teknologi Informasi .
4. Bapak Mhd. Basri, S.Si, M.Kom Sekretaris Program Studi Teknologi Informasi.
5. Ibu Yoshida Sary, S.Kom, S.E, M.Kom selaku pembimbing skripsi yang telah memberikan arahan, saran, dan bimbingan dengan penuh kesabaran dan ketulusan.

6. Ibunda Yudhy Rahmi Putri dan Anisha Syafira Putri sebagai saudara kandung saya yang telah memberikan dukungan dari segi moral maupun materiil serta menjadi motivasi penulis untuk dapat menyelesaikan tugas akhir ini.
7. Seluruh Bapak dan Ibu Dosen di Jurusan Teknologi Informasi Fakultas Ilmu Komputer & Teknologi Informasi Universitas Muhammdiyah Sumatera Utara atas ilmu dan didikannya selama perkuliahan.
8. Sahabat saya Tamam Auliya Habib, dan Angga Tifansyah Siregar yang telah membantu dan mendukung penulis dalam menyelesaikan tugas akhir ini.
9. Teman-teman seperjuangan jurusan teknologi informasi UMSU.
10. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang telah membantu dalam penyelesaian skripsi ini.

Penulis menyadari bahwa dalam penulisan skripsi ini masih terdapat kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa yang akan datang. Akhir kata, penulis berharap skripsi ini dapat bermanfaat bagi semua pihak yang berkepentingan, khususnya bagi perkembangan ilmu pengetahuan di bidang Nama Bidang Ilmu.

Medan, 16 September 2025

Penulis

M. Ichwan Gifari

“PENERAPAN METODE ALGORITMA EXTREME GRADIENT BOOSTING DALAM MEMPREDIKSI PENJUALAN THRIFTING PADA TOKO CCHASE”

ABSTRAK

Thrifting di zaman sekarang sangat mudah dijangkau karena sudah banyak bermunculan toko thrift di perkotaan. Namun permasalahan yang sering terjadi pada toko thrifting ialah menghadapi masalah stok dan prediksi penjualan. Seperti pada toko Cchase, penentuan jumlah stok sering tidak tepat karena variasi barang thrifting yang beragam dan pasokan yang tidak menentu, sehingga menimbulkan risiko penumpukan barang yang tidak terjual. Untuk mengatasi hal ini, penelitian menggunakan teknologi *machine learning* dengan algoritma Extreme Gradient Boosting (XGBoost). Metode ini dipilih karena terbukti mampu memberikan hasil prediksi yang akurat pada data dengan pola yang kompleks. Data yang digunakan berupa catatan penjualan selama satu tahun terakhir yang kemudian diproses, dibagi menjadi data latih dan data uji, lalu diuji menggunakan metrik evaluasi Root Mean Squared Error (RMSE) dan Mean Absolute Error (MAE). Hasil penelitian menunjukkan nilai RMSE sebesar 0,8105 dan MAE sebesar 0,6643, yang berarti model cukup baik dalam memprediksi penjualan. Dari hasil prediksi untuk periode satu bulan ke depan, diperoleh kategori dengan penjualan tertinggi yaitu crewneck, hoodie, dan kaos. Dengan adanya hasil ini, diharapkan dapat membantu pelaku usaha thrift dalam mengelola stok lebih efisien serta menyusun strategi penjualan yang lebih tepat sasaran.

Kata Kunci: Thrifting, Prediksi Penjualan, *Machine learning*, XGBoost, RMSE, MAE.

“PENERAPAN METODE ALGORITMA EXTREME GRADIENT BOOSTING DALAM MEMPREDIKSI PENJUALAN THRIFTING PADA TOKO CCHASE”

ABSTRACT

Thrifting nowadays is very easy to access since thrift stores have been rapidly emerging in urban areas. However, one of the common problems faced by thrift shops is related to stock management and sales prediction. For example, at Cchase store, stock estimation is often inaccurate due to the wide variety of thrift items and the uncertain supply, which leads to the risk of unsold inventory. To overcome this issue, this study applies *machine learning* technology using the Extreme Gradient Boosting (XGBoost) algorithm. This method was chosen because it has been proven to provide accurate prediction results on data with complex patterns. The data used consist of one year of sales records, which were processed, split into training and testing sets, and then evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics. The results show that the model achieved an RMSE of 0.8105 and an MAE of 0.6643, indicating that the model performs well in predicting sales. Furthermore, the prediction results for the upcoming month reveal the top three product categories with the highest sales, namely crewnecks, hoodies, and t-shirts. These findings are expected to help thrift business owners manage stock more efficiently and develop more effective sales strategies

Keywords: Thrifting, Sales Prediction, *Machine learning*, XGBoost, RMSE, MAE

DAFTAR ISI

LEMBAR PENGESAHAN ..... **i**
PERNYATAAN ORISINALITAS ..... **ii**
PERNYATAAN PERSETUJUAN PUBLIKASI ..... **iii**
RIWAYAT HIDUP ..... **iv**
KATA PENGANTAR ..... **v**
ABSTRAK ..... **vii**
ABSTRACT ..... **viii**
DAFTAR ISI ..... **ix**
DAFTAR TABLE ..... **xi**
DAFTAR GAMBAR ..... **xii**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang Masalah 1
1.2 Rumusan Masalah 4
1.3 Batasan Masalah 4
1.4 Tujuan Penelitian 5
1.5 Manfaat Penelitian 5
BAB II LANDASAN TEORI **6**
2.1 *Thrifting* 6
2.2 *Machine learning* 7
2.3 XGBoost 7
2.4 Model Evaluasi 8
2.5 *Python* 9
2.6 Penelitian Terkait 10
BAB III METODOLOGI PENELITIAN **20**
3.1 Pendeketan Penelitian 20
3.2 Pengumpulan Data 20
3.3 Tahapan Penelitian 22
3.4 Alur Penelitian 24
3.5 Analisis Perhitungan 26
BAB IV HASIL DAN PEMBAHASAN **28**
4.1. Hasil Pengumpulan Data 28
4.2. Import Library 29
4.3. Import Dataset 30
4.4. Pre-Procesing Data 30
4.5. Data Preparation 32
4.6 Modelling XGBoost 33
4.7 Evaluasi Model 34
4.8 Hasil Prediksi 35

4.9 Visualisasi Prediksi..... 38
BAB V PENUTUP.....41
5.1 Kesimpulan.....41
5.2 Saran.....42
DAFTAR PUSTAKA.....43

DAFTAR TABLE

Table 2.1 Penelitian Terkait..... 16

Table 3.1 Kriteria RMSE dan MAE 25

DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian 22

Gambar 3.2 Alur Penelitian 24

Gambar 3.3 Prediksi Perhitungan 26

Gambar 4.1 Hasil Pengumpulan Data 28

Gambar 4.2 Bar Chart Prediksi Penjualan 39

Gambar 4.3 Line Chart Perbandingan Nilai Aktual dan Prediksi 40

BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Thrifting adalah praktik membeli barang bekas, terutama pakaian milik orang lain, lalu menjualnya kembali. Karena sebagian orang menyukai pakaian bekas yang unik dan terbatas, serta sesuai dengan tren mode terkini yang lebih condong ke estetika vintage atau retro, pakaian bekas semakin populer (Jury, 2024). Berbeda dengan pakaian baru, pakaian bekas diimpor dari luar negeri dan menampilkan merek serta gaya populer yang hanya dimiliki sedikit orang (Utarsih, 2024). Ada banyak cara untuk mendapatkan pakaian atau barang berkualitas tinggi lainnya dengan harga yang jauh lebih rendah daripada harga normal. Mengonsumsi pakaian bekas, juga dikenal sebagai "*Thrifting*", telah berkembang menjadi aspek utama budaya ritel global dalam beberapa tahun terakhir (Jury, 2024). Remaja masa kini yang ingin tampil bergaya dan kekinian tidak perlu pergi ke mal untuk mencapainya karena mereka dapat melakukannya di toko barang bekas, yang merupakan tempat yang menjual pakaian bekas (Adji, 2022).

Usaha *Thrifting* menyediakan berbagai macam barang yang beragam dan unik dalam jumlah terbatas *(limited quantity)* (Jury, 2024). Salah satu *Thrift shop* di kota Medan yang menyediakan produk *Thrifting* seperti toko *Cchase.* Bagi banyak pelaku usaha *Thrift* terutama yang beroperasi dalam skala kecil atau menengah, pakaian bekas merupakan mata pencaharian utama yang mendukung keberlangsungan usaha mereka (Jury, 2024)

Dalam rangka mengelola industri thrifting, pemerintah melakukan imbauan kepada masyarakat terkait importasi pakaian bekas. Misalnya, Peraturan Menteri Perdagangan (PERMENDAG) No. 40 Tahun 2022 menyatakan dalam pasal 2 ayat 3 bahwa tas bekas, karung bekas, dan pakaian bekas merupakan barang impor yang dilarang. Hanya sedikit pelaku usaha yang tidak menawarkan pakaian bekas impor di toko thrifting (Jury, 2024). Hanya importasi pakaian bekas ke Indonesia yang dilarang oleh pemerintah melalui Menteri Perdagangan. Berdasarkan berbagai Peraturan Menteri, dapat disimpulkan bahwa pelaku usaha Toko Thrifting dan Preloved wajib mendapatkan izin dari Menteri di bidang perdagangan. Selain itu, perdagangan pakaian bekas oleh Toko Thrifting dan Preloved tidak dianggap pelanggaran apabila pakaian bekas yang dijual oleh pelaku usaha tersebut berasal dari milik pribadi dan bukan merupakan hasil impor pakaian bekas (Santika, 2021). Beberapa alternatif strategi bisnis yang direkomendasikan termasuk membuka *Thrift shop* lokal yang memanfaatkan pakaian bekas dalam negeri melalui donasi, membuka toko dengan sistem konsinyasi untuk mengurangi kebutuhan modal awal, serta beralih ke model dropshipping atau reseller barang lokal yang lebih legal (Kirana, 2025). Maka dapat dipahami bahwa belum ada pengaturan yang mengatur secara khusus mengenai praktik penjualan barang bekas, khususnya pakaian bekas dalam negeri (Santika, 2021).

Pelaku usaha juga mengalami kesulitan dalam menghadapi ketidakpastian dalam pengelolaan stok dan prediksi permintaan pasar. Berbeda dengan produk retail baru, barang *Thrifting* memiliki variasi yang tinggi dan pasokan yang tidak

selalu ada. Jika salah dalam menentukan persediaan stok, maka akan berpengaruh terhadap keuangan atau penjualan pada toko karena produk *Thrift* tersebut lambat terjual (Widiana, 2023).

Dalam menghadapi permasalahan yang sering dihadapi oleh pelaku usaha dari toko *Cchase,* seperti ketidakpastian jumlah permintaan, kesulitan mengelola stok barang yang beragam, serta risiko penumpukan produk yang tidak terjual penerapan teknologi seperti *machine learning* dapat membantu dalam memprediksi penjualan dan pengelolaan produk *Thrifting*.

Telah terbukti bahwa pelaku bisnis dapat menghasilkan proyeksi penjualan yang lebih akurat, efisien, dan adaptif terhadap pasar dengan memanfaatkan teknologi *Machine learning* (Riza, 2022). Oleh karena itu, studi ini menggunakan metode *Extreme Gradient Boosting* (XGBoost) dalam pendekatan prediktif berbasis *Machine learning*. Pendekatan ini telah menunjukkan hasil yang menjanjikan dalam sejumlah aplikasi, seperti klasifikasi dan prediksi data (Pratama, 2024). XGBoost adalah sistem prediksi penjualan untuk barang dagangan toko barang bekas yang sangat bervariasi karena kemampuannya memproses data dalam jumlah besar dan mengidentifikasi pola yang rumit. Harga, kategori produk, dan waktu penjualan hanyalah beberapa contoh korelasi non-linier dan kompleks yang dapat ditangkap oleh XGBoost dalam konteks prediksi penjualan, menghasilkan prediksi yang lebih akurat dan fleksibel dalam merespons perubahan tren pasar (Pratama, 2024).

Dengan adanya penelitian ini dapat memberikan solusi dan wawasan serta menambah strategi bisnis kepada pelaku bisnis di toko *Cchase* agar memahami dan memanfaatkan teknologi menggunakan metode *machine learning* khususnya dalam

algoritma XGBoost untuk menganalisa dan memprediksi penjualan *Thrift* agar lebih efisien dan akurat dalam menentukannya. Ini juga dapat menjadi solusi dalam menentukan persediaan jenis pakaian ataupun barang yang dapat dipasok untuk dijadikan stok buat penjualan agar terhindar dari berbagai masalah yang dapat membuat perputaran produk menjadi sangat lambat.

1.2 Rumusan Masalah

1. Bagaimana penerapan *machine learning* dengan algoritma *Extreme Gradient Boosting* (XGBoost) ini dalam memprediksi penjualan produk *thrift* bulan depan agar dalam mengelola stok terhindar dari penimbunan produk yang berlebih.
2. Sejauh mana metrik evaluasi *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE) dapat digunakan untuk mengukur kinerja prediksi penjualan produk Thrifting yang dihasilkan oleh algoritma XGBoost.

1.3 Batasan Masalah

1. Penelitian ini hanya akan menggunakan *machine learning* dengan metode algoritma *Extreme Gradient Boosting* (XGBoost), pada sistem pengklasifikasinya.
2. Penelitian ini tidak mencakup faktor tentang dampak kesehatan dan pengaruh terhadap lingkungan.
3. Penelitian ini hanya akan mengambil data historis penjualan produk *thrift* 1 tahun terakhir.
4. Penelitian ini hanya mengambil tempat penelitian yang dilakukan pada toko *Cchase* Medan.

1.4 Tujuan Penelitian

1. Mengetahui cara kerja penerapan metode algoritma *Extreme Gradient Boosting* (XGBoost) dalam memprediksi penjualan produk *Thrifting* yang akan tren dan diminati oleh konsumen pada bulan depan.
2. Mengevaluasi akurasi model prediksi penjualan dengan menggunakan metrik evaluasi seperti *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE), untuk menilai sejauh mana model mampu memberikan hasil prediksi yang mendekati aslinya.

1.5 Manfaat Penelitian

1. Menjadi referensi bagi penelitian selanjutnya dalam penggunaan *machine learning* dengan metode algoritma (XGBoost) dalam hal menganalisa dan memprediksi penjualan produk *Thrifting.*
2. Bagi mahasiswa yaitu memberikan wawasan mendalam mengenai penerapan algoritma XGBoost, dalam hal memprediksi penjualan usaha atau bisnis lokal.
3. Bagi kampus yaitu dapat menjadi bahan evaluasi dan inspirasi dalam melakukan penelitian kampus selanjutnya.
4. Bagi perusahaan yaitu memberikan manfaat untuk mendorong perusahaan atau pelaku usaha *Thrifting* agar perputaran penjualan produk menjadi tidak lambat karena dapat berpengaruh terhadap pendapatan toko dan juga perilaku atau kepuasan konsumen.
5. Bagi masyarakat memberikan kesadaran kepada masyarakat akan pentingnya penggunaan teknologi dalam pengelolaan usaha lokal yang berkelanjut

BAB II
LANDASAN TEORI

2.1 *Thrifting*

Istilah *thrifting* adalah sebutan bagi seseorang yang sedang membeli barang atau pakaian bekas. Pakaian bekas sangat populer di kalangan sebagian orang, terutama remaja. Tentu saja, ini semua berkaitan dengan gaya hidup zaman yang semakin berkembang dengan mengikuti tren, dan membuat tren yang digemari remaja. Selain itu, sebagian remaja mempertimbangkan alasan mereka menggunakan pakaian bekas, karena harganya yang relatif murah dan pakaian bekas ini memiliki ciri khas tersendiri, seperti edisi terbatas dan masih layak pakai (Nadhila, 2023). Dalam konteks ini, thrifting tidak hanya dipahami sebagai aktivitas membeli barang bekas, tetapi konsumen dapat merasakan sensasi dari pakaian yang dibelinya, terutama jika mendapatkan barang langka dengan harga lebih murah.

Produk *thrifting* juga memiliki berbagai macam seperti pakaian yaitu baju kaos, celana, kemeja, hoodie, dan lain sebagainya yang juga memiliki beragam brand. Seiring dengan perkembangan zaman, aktivitas berbelanja *thrift* sekarang dapat dijangkau dengan mudah karena adanya aplikasi berbelanja *online*, serta semakin menjamur *thrift shop* di berbagai daerah (Adji, 2022). Dengan semakin banyaknya toko *thrift shop*, konsumen tidak perlu repot lagi mecari barang atau pakaian bekas dengan waktu yang lama dipasar loak. Hal ini terkait langsung dengan alasan orang membeli barang dari toko barang bekas saat ini, yaitu untuk menjaga gaya berpakaian dan mengikuti perkembangan zaman sambil juga memperhitungkan biaya dan penghematan.

2.2 *Machine learning*

Machine learning merupakan cabang ilmu komputer yang berfokus pada pengembangan sistem yang dapat belajar secara mandiri tanpa perlu pemrograman manusia berulang kali. Namun, sebelum menghasilkan data tentang perilaku objek, pembelajaran mesin membutuhkan data awal untuk dipelajari (Saputro, 2024). Untuk memprediksi penjualan produk barang bekas, teknologi ini menganalisis data penjualan dalam jumlah besar, memberikan wawasan yang sebelumnya tidak tersedia melalui metode tradisional.

Dalam *thrifting*, metode *machine learning* digunakan untuk meningkatkan penjualan barang atau pakaian bekas agar terhindar dari timbunan stok yang perputarannya produk atau keuangan menjadi sangat lambat. Pelaku usaha menggunakan metode *machine learning* untuk menganalisis data penjualan, memantau tren yang akan datang, dan mengetahui perilaku atau kepuasan konsumen. Pendekatan berbasis data ini membantu pelaku usaha mengembangkan prediksi penjualan yang lebih efektif.

2.3 XGBoost

Metode XGBoost merupakan pengembangan dari algoritma penguat pohon gradien berbasis *ensemble* yang berhasil mengatasi permasalahan pembelajaran mesin yang kompleks (Yulianti, 2022). Evolusi lain dari teknik Penguatan Gradien, sebuah pendekatan ensemble untuk model berbasis pohon keputusan, adalah XGBoost. Untuk klasifikasi lemah, XGBoost menggabungkan beberapa model pembelajaran mesin. Dengan menggunakan pelatihan inkremental, kombinasi model ini bertujuan untuk meningkatkan kinerja.

Hasil klasifikasi dari model *machine learning* sebelumnya yang disebut sebagai residuals atau *error* dijadikan bahan evaluasi agar kinerja algoritma dengan metode *ensemble* ini meningkat. XGBoost dapat mengoptimalkan modelnya untuk mencapai keseimbangan yang tepat antar presisi dan kompleks, yang membantu mencegah model dari *overfitting* pada data pelatihan (Yulianti, 2022). Fokus utama dalam algoritma XGBoost adalah pada penurunan nilai fungsi objek teratur, yang didefinisikan sebagai berikut:

$$\text{Obj}(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Keterangan**:**

θ = parameter model

n = Jumlah total data pelatihan

$l(y_i, \hat{y}_i)$ = Fungsi loss antara nilai aktual y_i dan prediksi $\hat{y}_i$

$\Omega(f_k)$ = Fungsi regularisasi terhadap model prediksi ke-k

2.4 Model Evaluasi

Untuk mengukur hasil performa dari algoritma yang digunakan untuk sebuah peramalan. Terdapat berbagai jenis metrik evaluasi peramalan yang sering diterapkan, namun metode yang paling umum digunakan meliputi *Root Mean Squared Error* (RMSE), dan *Mean Absolute Error* (MAE) (Nurak, 2025).

- *Root Mean Squared Error* (RMSE)

  Metode evaluasi RMSE merupakan akar kuadrat dari rata-rata selisih kuadrat antara nilai prediksi model dan nilai aktual dari data yang diamati. Nilai RMSE yang lebih kecil menunjukkan bahwa model memiliki tingkat prediksi yang lebih baik. Adapun rumus RMSE adalah sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2}$$

Keterangan:

$\hat{y_i}$ = nilai prediksi/model pada data ke-i

y_i = nilai aktual pada data ke-i

n = jumlah total data

- *Mean Absolute Error* (MAE)

  Metrik evaluasi MAE menghitung rata-rata dari selisih absolut antara nilai aktual dan nilai prediksi yang dihasilkan oleh model, MAE memberikan bobot yang sama untuk setiap kesalahan tanpa memperhatikan arah (positif atau negatif) dari kesalahan tersebut. Adapun rumus MAE adalah sebagai berikut:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i|$$

Keterangan:

n = jumlah total data

y_i = nilai aktual pada data ke-i

$\hat{y_i}$ = nilai prediksi/model pada data ke-i

$|y_i - \hat{y_i}|$ = nilai absolut dari selisih antara nilai aktual dan nilai prediksi

2.5 *Python*

Python adalah bahasa pemrograman tingkat tinggi yang bersifat *interpreted*, *multi-paradigm*, dan *open-source*, pertama kali dikembangkan oleh Guido van Rossum dan dirilis secara resmi pada tahun 1991. *Python* juga merupakan bahasa pemrograman yang multi fungsi salah satunya pada bidang *Machine learning* dan

Deep Learning. *Python* dipilih untuk studi ini karena sintaksisnya yang ramah pengguna, pustaka yang luas, dan dukungan komunitas yang kuat karena bersifat sumber terbuka. Anda dapat menggunakan IDE seperti Visual Studio Code, Sublime Text, atau PyCharm untuk menulis kode sumber Python, atau Anda dapat memanfaatkan IDE daring seperti Google Colab dan Jupyter Notebook (Alfarizi, 2023).

2.6 Penelitian Terkait

Penelitian relevan (Wijaya, 2024) yang berjudul "*Meningkatkan Prediksi Penjualan Retail Xyz Dengan Teknik Optimasi Random Search Pada Model XGBoost*" Penelitian ini membahas tentang metode baru untuk prediksi penjualan dengan menerapkan XGBoost, yang dioptimalkan melalui perbandingan tiga metode optimasi: pencarian acak (*random search*), pencarian grid (*grid search*), dan optimasi *Bayesian*. Penelitian ini mengadaptasi struktur waktu seperti yang digunakan dalam studi Li dan rekan, merujuk pada dataset Amazon UK Sales Forecasting 2019–2021 untuk validasi kinerja model prediktif berbasis XGBoost. Tujuannya adalah untuk mendapatkan analisis perbandingan terbaik dan meningkatkan akurasi prediksi. Kebaruan dari model ini terletak pada penentuan nilai optimal untuk setiap metode optimasi yang diterapkan pada XGBoost. Hasil evaluasi menunjukkan bahwa teknik optimasi *grid search* memberikan performa terbaik pada model XGBoost, dengan peningkatan nilai evaluasi R^2 dari 97,31 menjadi 98,41. Analisis model yang diusulkan ini dapat membantu pemilik bisnis ritel dalam melakukan prediksi penjualan secara lebih akurat untuk mendukung perkembangan proses bisnis.

Pada penelitian (Nurmaya, 2024) yang berjudul *"Prediksi Harga Saham pada Bank BCA Tbk dengan Teknik XGBoost (Extreme Gradient Boosting)"* merupakan kajian kuantitatif yang bertujuan untuk meramalkan harga penutupan saham Bank BCA Tbk dalam rentang waktu lima tahun, mulai dari Januari 2019 hingga Desember 2023. Salah satu kekuatan utama studi ini terletak pada proses rekayasa fitur yang mengintegrasikan indikator teknikal seperti EMA (*Exponential Moving Average*), SMA (*Simple Moving Average*), RSI (*Relative Strength Index*), dan MACD (*Moving Average Convergence Divergence*). Melalui pendekatan ini, model berhasil memberikan hasil prediksi yang mendekati nilai aktual dengan nilai MAPE sebesar 2,48%, yang menunjukkan tingkat akurasi sangat tinggi. Hasil ini menunjukkan bahwa XGBoost mampu menangani kompleksitas pola harga saham dan beradaptasi terhadap dinamika pasar yang tidak stabil.

Pada penelitian (Winurputra, 2025) yang berjudul *"Peramalan Penjualan Produk Menggunakan Extreme Gradient Boosting (XGBoost) dan Kerangka Kerja CRISP-DM untuk Pengoptimalan Manajemen Persediaan (Studi Kasus: UB Mart)"* Penelitian ini memanfaatkan algoritma XGBoost sebagai model prediktif utama, yang dikombinasikan dengan tahapan *Cross-Industry Standard Process for Data Mining (*CRISP-DM*)* untuk memastikan pendekatan yang sistematis dan terstruktur. Penelitian ini meliputi catatan penjualan dari Januari 2020 hingga Juni 2023 dalam studi kasus UB Mart. Hasil evaluasi menunjukkan bahwa skenario ketiga yang menggabungkan fitur deret waktu dan fitur penanda seperti jumlah transaksi harian dan transaksi dalam volume besar menghasilkan performa terbaik dengan nilai MAE rata-rata sebesar 11,58 dan RMSE sebesar 29,19**,** mengindikasikan bahwa model mampu memberikan hasil prediksi yang cukup

akurat. Dengan demikian, studi ini menunjukkan bahwa pendekatan berbasis *data science* melalui XGBoost dan CRISP-DM tidak hanya meningkatkan keandalan prediksi, tetapi juga memberikan kontribusi nyata dalam pengambilan keputusan operasional untuk pengelolaan persediaan di sektor ritel.

Pada penelitian oleh (Tiastama, 2024) yang berjudul *"Perbandingan Random Search dan Algoritma Genetika dalam Penyetelan Hyperparameter XGBoost pada Retail Sales Forecasting"* mengkaji upaya peningkatan akurasi peramalan penjualan ritel melalui optimasi *hyperparameter* algoritma XGBoost. Fokus utama penelitian ini terletak pada perbandingan dua teknik optimasi yaitu *Random Search* dan Algoritma Genetika dalam menyetel *hyperparameter* XGBoost, yang terbukti sangat berpengaruh terhadap performa model. Berdasarkan temuan studi yang menggunakan algoritma evolusioner dan random search pada XGBoost untuk melakukan penyetelan hiperparameter menggunakan dataset perusahaan ritel Rossmann Store, yang mencakup 168.868 data pengujian dan 844.338 data pelatihan. Dengan nilai RMSE 0,123 dan 0,122 untuk fase pelatihan dan pengujian, dapat dikatakan bahwa pendekatan pencarian acak berkinerja lebih baik daripada algoritma genetika. Sebaliknya, nilai RMSE algoritma genetika pada fase pelatihan dan pengujian masing-masing adalah 0,333 dan 0,332. Menemukan hiperparameter terbaik pada XGBoost untuk peramalan penjualan ritel dapat dilakukan menggunakan pencarian acak dan algoritma genetika, karena perbedaan nilai RMSE tidak signifikan.

Pada Penelitian oleh (Riza, 2022) yang berjudul *"Analisis dan Prediksi Data Penjualan Menggunakan Machine learning dengan Pendekatan Ilmu Data"* merupakan studi komprehensif yang memanfaatkan kerangka kerja ilmu data untuk

meramalkan penjualan ritel masa depan secara akurat. Data set yang dikumpulkan dalam penelitian ini ialah Big mart sales dengan 10 toko yang terletak dilokasi berbeda dengan masing-masing toko memiliki 1559 produk berbeda sesuai pengumpulan data tahun 2013 dengan atribut tertentu dari setiap produk dan toko telah ditentukan. Penelitian ini mengaplikasikan dan membandingkan lima algoritma *machine learning*, yaitu *Lasso Regressor, Linear Regression, Random Fores*t, LightGBM, dan XGBoost, dengan tujuan mencari model prediksi terbaik berdasarkan metrik evaluasi MAE, RMSE, dan R^2. Hasil evaluasi menunjukkan bahwa XGBoost memperoleh performa terbaik dengan skor R^2 sebesar 0.6142, diikuti oleh LightGBM sebesar 0.6094, sementara *Random Forest* menunjukkan performa terlemah dengan R^2 sebesar 0.3268. Dengan demikian, penelitian ini menyimpulkan bahwa penggunaan algoritma XGBoost dalam kombinasi dengan tahapan sistematis CRISP-DM dapat menjadi strategi prediktif yang efektif dan akurat dalam pengambilan keputusan bisnis berbasis data.

Pada penelitian oleh (Anggara, 2025) yang berjudul *"Pengembangan Sistem Prediksi Harga dan Rekomendasi Mobil Bekas Berbasis Machine learning".* Data yang digunakan dalam penelitian ini diambil dari platform *Kaggle*, dengan total 6.019 data mobil bekas yang masing-masing memiliki 10 atribut utama, termasuk merek, tahun produksi, jarak tempuh, jenis bahan bakar, transmisi, tipe kepemilikan, hingga harga jual. Hasil pengujian menunjukkan bahwa XGBoost *Regressor* unggul signifikan dengan tingkat akurasi 97.79% pada data latih dan 89.90% pada data uji, serta nilai MSE sebesar 2.3129 dan RMSE sebesar 1.5208. Ini jauh lebih baik dibandingkan dengan regresi linear dan turunannya yang hanya mencapai akurasi sekitar 81.6% dan RMSE di atas 2.0. Keunggulan XGBoost

terletak pada kemampuannya menangani relasi non-linier antar fitur dan optimasi parameter melalui teknik *hyperparameter tuning*.

Pada penelitian oleh (Rayadin, 2024) yang berjudul *"Implementasi Ensemble Learning Metode XGBoost dan Random Forest untuk Prediksi Waktu Penggantian Baterai Aki"*. Penelitian ini mengangkat isu penting dalam dunia otomotif terkait prediksi waktu penggantian baterai aki pada kendaraan bermotor, khususnya mobil. Dataset yang digunakan dalam penelitian ini merupakan data historis penggantian baterai aki mobil Toyota dari salah satu perusahaan otomotif di Kota Makassar, mencakup rentang waktu selama 7 tahun dari tahun 2017 hingga 2023. Hasil evaluasi menunjukkan bahwa model XGBoost menghasilkan performa terbaik dari sisi MAPE (10.96%), MAE (38.76), dan RMSE (38.760). Temuan ini sangat relevan dalam konteks layanan otomotif modern yang membutuhkan efisiensi, kecepatan, dan ketepatan layanan.

Penelitian oleh (Siringoringo, 2021) yang berjudul *"Segmentasi Dan Peramalan Pasar Retail Menggunakan XGBoost Dan Principal Component Analysis".* Penelitian ini membahas penerapan metode *machine learning* untuk mendukung segmentasi pasar dan peramalan penjualan dalam sektor retail online yang berkembang pesat di Indonesia. Dataset yang digunakan dalam studi ini berasal dari UCI *Machine learning Repository*, yaitu online retail dataset dengan total 525.461 item data. Hasil pengujian menunjukkan bahwa model XGBoost mampu melakukan klasifikasi dengan akurasi rata-rata 88%**,** *precision* 78%, *recall* 78%, dan *F1-score* 78%. Selain itu, kurva ROC untuk XGBoost menunjukkan nilai tertinggi dibanding metode lain seperti *AdaBoost* dan *Gradient Boost*.

Pada penelitian (Amiruddin Bengnga, 2023) yang berjudul *"Penerapan Algoritma XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah".* Berdasarkan hasil pemilihan 12 atribut yang ada, temuan penelitian menunjukkan bahwa algoritma XGBoost dapat digunakan untuk memilih atribut terpenting dalam K-Means dalam mengelompokkan prioritas penerima bantuan KIP Kuliah. 3 atribut terpenting dipilih, yaitu pekerjaan ayah, pendapatan ibu. Luas Bangunan, dan nilai indeks terkecil, 0,819, diperoleh dari hasil evaluasi temuan pengelompokkan.

Penelitian oleh (Hidayat, 2025) yang berjudul *"Analisis Performa Algoritma XGBoost, GRU, dan Prophet dalam Peramalan Penjualan Obat untuk Optimasi Rantai Pasok Farmasi".* Dataset studi ini, yang mencakup 2.106 baris data penjualan obat harian dari sebuah perusahaan farmasi di Niš, Serbia, diperoleh dari platform Kaggle. Dengan MSE 16,1885, RMSE 4,0234, MAE 2,6427, MAPE 4,35%, dan R2 0,9646, hasil evaluasi kinerja menunjukkan bahwa metode XGBoost memberikan hasil terbaik, terutama pada rasio 60:40. Ini berarti model tersebut dapat memperhitungkan hampir 96% variasi dalam data. Di sisi lain, algoritma Prophet dan GRU tidak berkinerja sebaik itu. MSE GRU berada di antara 363 dan 406, dan R2-nya kurang dari 0,25.

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti (Tahun)	Penelitian	Hasil
1.	Hendra Wijaya, Dandy Pramana Hostiadi, Evi Triandini (2024)	Meningkatkan Prediksi Penjualan Retail Xyz Dengan Teknik Optimasi *Random Search* Pada Model XGBoost	Hasil penelitian menunjukkan bahwa teknik optimasi *grid search* memberikan performa terbaik pada model XGBoost, dengan peningkatan nilai evaluasi R^2 dari 97,31 menjadi 98,41. Analisis model yang diusulkan ini dapat membantu pemilik bisnis ritel dalam melakukan prediksi penjualan secara lebih kurat untuk mendukung perkembangan proses bisnis.
2.	Septia Tri Nurmaya, Budi Soesilo, Doni Abdul Fatah (2024)	Prediksi Harga Saham Pada Bank BCA Tbk dengan Teknik XGBoost (*Extreme Gradient Boosting*)	Hasil penelitian ini menunjukkan model berhasil memberikan hasil prediksi yang mendekati nilai aktual dengan nilai MAPE sebesar 2,48%, yang menunjukkan tingkat akurasi sangat tinggi. Hasil ini menunjukkan bahwa XGBoost mampu menangani kompleksitas pola harga saham dan beradaptasi terhadap dinamika pasar yang tidak stabil.
3.	Raihan Winurputra, Dian Eka Ratnawati (2025)	Peramalan Penjualan Produk Menggunakan *Extreme Gradient Boosting* (XGBoost) dan Kerangka Kerja CRISP-DM untuk Pengoptimalan Manajemen Persediaan (Studi Kasus: UB Mart)	Hasil penelitian menghasilkan performa terbaik dengan nilai MAE rata-rata sebesar 11,58 dan RMSE sebesar 29,19**,** mengindikasikan bahwa Model XGBoost yang dibangun dengan

			metodologi CRISP-DM memberikan hasil prediksi yang cukup akurat.
4.	Sheren Afryan Tiastama, Indra Budi (2024)	Perbandingan *Random Search* dan Algoritma Genetika dalam Penyetelan *Hyperparameter* XGBoost pada Retail *Sales Forecasting*	Dengan nilai RMSE masing-masing 0,123 dan 0,122 untuk fase pelatihan dan pengujian, studi ini menemukan bahwa pendekatan random search mengungguli algoritma genetika. Sementara itu, nilai RMSE untuk algoritma genetika adalah 0,333 dan 0,332. Hiperparameter terbaik untuk XGBoost dapat ditemukan menggunakan algoritma genetika atau random search.
5.	Ferdy Riza (2022)	Analisis dan Prediksi Data Penjualan Menggunakan *Machine learning* Dengan Pendekatan Ilmu Data	Hasil penelitian menunjukkan bahwa XGBoost memperoleh performa terbaik dengan skor R^2 sebesar 0.6142, diikuti oleh LightGBM sebesar 0.6094, sementara *Random Forest* menunjukkan performa terlemah dengan R^2 sebesar 0.3268.
6.	Jerry Anggara, Fauzi Rahmat Ramadhan, Achmad Fikri Syofian, Apriansa Arwandi Panjaitan, Raka Fazah Fithra (2025)	Pengembangan Sistem Prediksi Harga dan Rekomendasi Mobil Bekas Berbasis *Machine learning*	Hasil pengujian menunjukkan bahwa XGBoost *Regressor* unggul signifikan dengan tingkat akurasi 97.79% pada data latih dan 89.90% pada data uji, serta nilai MSE sebesar 2.3129 dan RMSE sebesar 1.5208. Ini jauh lebih baik dibandingkan dengan regresi linear dan turunannya.

7.	Muhamad Amhar Rayadin, Mustarum Musaruddin, Rizal Adi Saputra, Isnawaty (2024)	Implementasi Ensemble Learning Metode XGBoost dan Random Forest untuk Prediksi Waktu Penggantian Baterai Aki	Hasil evaluasi menunjukkan bahwa model XGBoost menghasilkan performa terbaik dari sisi MAPE (10.96%), MAE (38.76), dan RMSE (38.760).
8.	Rimbun Siringoringo, Resianta Perangin angin, Mufria J. Purba (2021)	Segmentasi dan Peramalan Pasar Ritel Menggunakan XGBoost dan Analisis Komponen Utama	Hasil pengujian menunjukkan bahwa model XGBoost mampu melakukan klasifikasi dengan akurasi rata-rata 88%**,** *precision* 78%, *recall* 78%, dan *F1-score* 78%. Selain itu, kurva ROC untuk XGBoost menunjukkan nilai tertinggi dibanding metode lain seperti *AdaBoost* dan *Gradient Boost*, yaitu nilai AUC sebesar 0.98, menandakan performa yang sangat baik dalam klasifikasi *multiclass*.
9.	Amiruddin Bengnga, Rezqiwati Ishak (2023)	Penerapan XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah	Hasil penelitian menyimpulkan bahwa dengan menerapkan algoritma XGBoost untuk memilih atribut terpenting dalam K-Means untuk pengelompokan prioritas penerima bantuan KIP Kuliah, tiga dari 12 atribut terpilih sebagai yang terpenting: (1) Pekerjaan Ayah, (2) Pendapatan Ibu, dan (3) Luas Bangunan. Berdasarkan hasil evaluasi hasil pengelompokan, nilai indeks terkecil adalah 0,819.

10.	Muhammad Taufik Hidayat, Mulia Sulistiyono (2025)	Analisis Performa Algoritma XGBoost, GRU, dan Prophet dalam Peramalan Penjualan Obat untuk Optimasi Rantai Pasok Farmasi	Hasil evaluasi performa menunjukkan bahwa algoritma XGBoost memberikan hasil terbaik, terutama pada rasio 60:40, dengan MSE sebesar 16.1885**,** RMSE 4.0234**,** MAE 2.6427**,** MAPE 4.35%, dan R^2 sebesar 0.9646, menunjukkan bahwa model mampu menjelaskan lebih dari 96% variasi dalam data. Sebaliknya, algoritma GRU dan Prophet menunjukkan performa yang lebih rendah. GRU mencatat MSE sebesar 363–406 dengan R^2 di bawah 0.25, sementara Prophet justru menghasilkan nilai R-squared negatif, yang mengindikasikan performa prediksi yang sangat buruk pada dataset ini.

BAB III
METODOLOGI PENELITIAN

3.1 Pendeketan Penelitian

Dalam penelitian ini, pendekatan yang digunakan adalah pendekatan prediktif dengan menggunakan model algoritma XGBoost untuk memprediksi penjualan produk *thrifting* pada toko *Cchase*. Algoritma ini dipilih karena memiliki kemampuan yang unggul dalam menangani dengan kompleksitas tinggi. Metode ini dipilih karena kemampuannya dalam mencegah *overfitting*, menangani data yang tidak terstruktur dan memiliki *missing value* juga menghasilkan prediksi yang cepat dan akurat pada data besar.

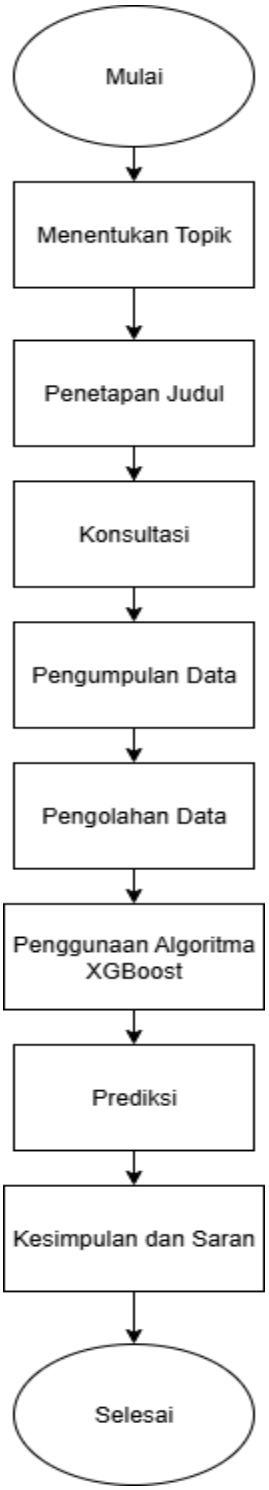
3.2 Pengumpulan Data

Dalam penelitian ini pengumpulan data dilakukan melalui dua pendekatan utama, yaitu wawancara semi terstruktur dan dokumentasi. wawancara dilakukan dengan narasumber yang berupa pemilik dari toko *Cchase* untuk memahami lebih dalam mengenai pola operasional penjualan dan faktor yang memengaruhi naik turunnya penjualan. Wawancara ini juga menghasilkan informasi mengenai permasalahan terkait kondisi *overstock* yang sering terjadi pada toko *Cchase*. Dokumentasi dilakukan untuk mendapatkan data historis penjualan harian produk *thrift* pada toko *Cchase*, dan memasukkannya ke Microsoft Excel. Data yang dikumpulkan seperti:

1. Tanggal
2. Hari
3. Minggu Penjualan
4. Bulan Penjualan
5. Merk

6. Kategori Produk
7. Harga
8. Diskon
9. Harga Akhir
10. Size
11. Kondisi
12. Metode Pembayaran
13. Total Penjualan Produk
14. Total Penjualan Bulanan

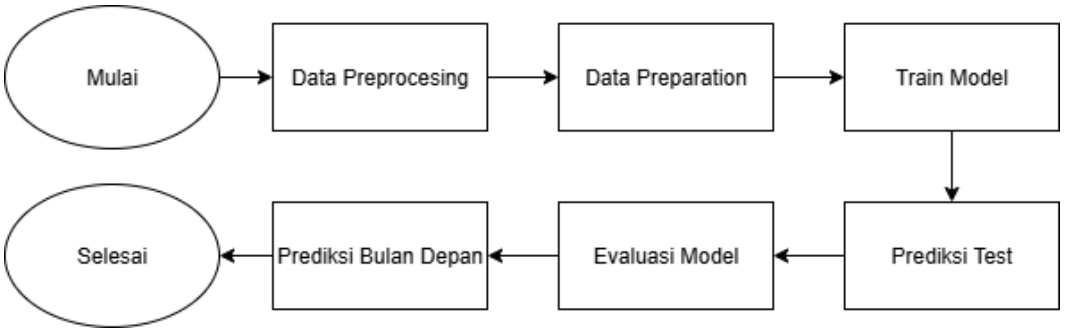
3.3 Tahapan Penelitian



Gambar 3.1 Tahapan Penelitian

1. Tahapan pertama menentukan topik penelitian yang akan diambil. Dimana penelitian ini mengambil topik *Machine learning* dan mencari permasalah yang bertujuan menentukan arah penelitian yang akan dilakukan.
2. Setelah itu penetapan judul, Dimana peneliti mengajukan judul penelitian dan disetuji oleh ketua program studi.
3. Konsultasi dimana setelah judul telah ditetapkan kemudian peneliti berdiskusi dengan dosen pembimbing yang telah ditetapkan terkait penelitian.
4. Pengumpulan data. Dalam tahap ini, data historis penjualan dari toko Cchase dikumpulkan selama setahun terakhir.
5. Setelah data terkumpul, tahap berikutnya adalah pengolahan data, yaitu data yang telah dipisahkan menjadi x (fitur), dan y (target) dengan rasio 80:20.
6. Tahap selanjutnya adalah penggunaan model algoritma. Dalam tahap ini, algoritma *Extreme Gradient Boosing* (XGBoost) digunakan untuk mempelajari pola-pola yang terdapat dalam data penjualan sebelumnya.
7. Proses berlanjut ke tahap prediksi. Di tahap ini, model digunakan untuk memproyeksikan penjualan produk dalam waktu ke depan berdasarkan pola historis yang telah dipelajari.
8. Setelah semuanya selesai lanjut kesimpulan dan saran. Kesimpulan diambil dari keseluruhan penelitian dan memberikan saran untuk penelitian selanjutnya.

3.4 Alur Penelitian



Gambar 3.2 Alur Penelitian

1. Diawali dengan tahap data preprocessing, yaitu proses pembersihan dan transformasi data mentah agar dapat digunakan dalam pemodelan. Pada tahap ini, data dicek untuk memastikan tidak ada nilai kosong, duplikasi, atau kesalahan format. Selain itu, seluruh variabel kategorikal seperti kategori, merk, hari, metode pembayaran, ukuran, dan kondisi produk diubah menjadi bentuk numerik menggunakan LabelEncoder. Transformasi ini penting karena algoritma *machine learning* hanya dapat bekerja dengan data numerik, sehingga semua informasi kategorikal perlu dikonversi terlebih dahulu.
2. Setelah proses preprocessing selesai, dilanjutkan dengan tahap data preparation. Pada tahap ini ditentukan variabel yang digunakan sebagai fitur (X) dan variabel target (y). Fitur merupakan sekumpulan variabel input yang memengaruhi hasil prediksi, seperti minggu ke, kategori produk, harga, harga akhir, diskon, dan metode pembayaran. Sedangkan target adalah total penjualan yang ingin diprediksi. Data kemudian dibagi menjadi dua bagian, yaitu data latih (training) dan data uji (testing). Pembagian ini dilakukan berdasarkan waktu, di mana seluruh data kecuali minggu terakhir digunakan sebagai data latih, sementara minggu terakhir dijadikan data uji.

3. Tahap selanjutnya adalah train model, yaitu melatih algoritma menggunakan data latih. Model yang digunakan adalah XGBoost, salah satu algoritma berbasis boosting yang efektif dalam menangani data tabular dan dikenal mampu memberikan prediksi dengan akurasi tinggi. Model ini dilatih dengan parameter tertentu seperti jumlah pohon keputusan, kedalaman pohon, dan laju pembelajaran. Selama proses pelatihan, model belajar mengenali pola hubungan antara variabel input dengan target sehingga mampu memprediksi jumlah penjualan berdasarkan fitur yang diberikan.
4. Setelah model selesai dilatih, dilakukan tahap prediksi pada data uji, yaitu data dari minggu terakhir. Pada tahap ini, model menghasilkan estimasi penjualan yang kemudian dibandingkan dengan data aktual. Perbandingan ini menjadi dasar untuk melakukan evaluasi model. Evaluasi dilakukan dengan menghitung nilai RMSE (*Root Mean Squared Error*) dan MAE (*Mean Absolute Error*). RMSE memberikan gambaran rata-rata kesalahan prediksi dengan penekanan lebih besar pada error yang tinggi, sedangkan MAE menunjukkan rata-rata kesalahan absolut yang lebih mudah dipahami karena langsung dalam satuan penjualan.

RMSE dan MAE	Pengertian
<10%	Kemampuan prediksi sangat baik
10%-20%	Kemampuan prediksi baik
20%-50%	Kemampuan prediksi layak
>50%	Kemampuan prediksi buruk

3.1 Table Kriteria RMSE dan MAE

5. Jika hasil evaluasi menunjukkan model bekerja dengan cukup baik, maka tahap selanjutnya adalah prediksi bulan depan. Untuk melakukannya, data minggu terakhir disalin lalu nilai minggu ke ditambah empat, sehingga dapat mewakili prediksi empat minggu ke depan. Model kemudian digunakan untuk menghasilkan prediksi penjualan di periode tersebut. Hasil prediksi selanjutnya diolah dengan cara mengelompokkan berdasarkan kategori produk, lalu dihitung total penjualan per kategori. Dari hasil agregasi ini, diperoleh tiga kategori dengan penjualan tertinggi yang diperkirakan akan laku di bulan mendatang.

3.5 Analisis Perhitungan

Untuk menguji performa awal model, dilakukan pelatihan model XGBoost menggunakan 5 sampel data dari transaksi penjualan produk *thrifting*. Variabel target yang diprediksi adalah Total Penjualan, yang diperoleh dari rumus:

$$\text{Total Penjualan Aktual} = \text{Harga Akhir} \times \text{Total Produk Terjual}$$

Tanggal	Hari	Minggu Ke	Bulan Penjualan	Merk	Kategori	Harga	Diskon (%)	Harga Akhir	Size	Kondisi	Metode Pembayaran	Total Penjualan Produk	Penjualan Aktual	MAE	RMSE
12/05/2024	Minggu	2	5	Uniqlo	Jeans	90000	0	90000	L	Bagus	E-Wallet	2	180000	0,02	0,00024
12/05/2024	Minggu	2	5	Adidas	Jacket	110000	10	99000	M	Bagus	Transfer Bank	3	297000	0,09	0,00879
12/05/2024	Minggu	2	5	H&M	kaos	90000	0	90000	S	layak	Cash	2	180000	0,02	0,00024
12/05/2024	Minggu	2	5	Levi's	kaos	70000	10	63000	L	layak	Cash	3	189000	0,02	0,00024
12/05/2024	Minggu	2	5	Levi's	jeans	100000	10	90000	L	Bagus	Transfer Bank	1	90000	0,01	0,00006

Gamar 3.3 Prediksi Perhitungan

Untuk mengevaluasi keakuratan prediksi, digunakan dua metrik utama, yaitu:

- *Mean Absolute Error* (MAE)

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}| = \frac{0.02+0,09+0,02+0,02+0,01}{5} = 0.03$$

Nilai *Mean Absolute Error* (MAE) mengindikasikan rata-rata selisih absolut antara nilai aktual dan prediksi. Dalam konteks ini, nilai MAE yang sangat kecil

menunjukkan bahwa secara umum, model menghasilkan prediksi yang sangat dekat dengan nilai sesungguhnya.

- *Root Mean Squared Error* (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2} = \frac{\sqrt{0.00024+0,00879+0,00024+0,00024+0,00006}}{5} = 0,044$$

Nilai *Root Mean Squared Error* (RMSE) yang lebih sensitif terhadap error besar, juga menunjukkan performa model yang sangat stabil, hal ini menandakan bahwa tidak terdapat error (outlier) dalam hasil prediksi.

BAB IV
HASIL DAN PEMBAHASAN

4.1. Hasil Pengumpulan Data

Tahap awal pada penelitian ini dimulai dengan pengumpulan data melalui wawancara dengan narasumber yaitu pemilik toko Cchase tersebut. Kemudian setelah melakukan wawancara, peneliti mengambil data penjualan toko *thrift* yang dikumpulkan dalam rentang waktu satu tahun kebelakang dimulai dari Mei 2024 hingga Mei 2025. Proses pengumpulan data disimpan dalam bentuk file excel untuk memudahkan dalam pengelolaan data. Data yang terkumpul mencakup tanggal penjualan, kategori produk, harga, total penjualan produk dan informasi terkait penjualan *thrifting* lainnya.

	Tanggal	Hari	Minggu Ke	Bulan Penjualan	Merek	Kategori	Harga	Diskon (%)	Harga Akhir	Size	Kondisi	Metode Pembayaran	Total Penjualan	Total Penjualan Bulanan
2	12/05/2024	Minggu	2	5	Uniqlo	Jeans	90000	0	90000	L	Bagus	E-Wallet	2	425
3	12/05/2024	Minggu	2	5	Adidas	Jacket	110000	10	99000	M	Bagus	Transfer Bank	3	425
4	12/05/2024	Minggu	2	5	H&M	kaos	90000	0	90000	S	layak	Cash	2	425
5	12/05/2024	Minggu	2	5	Levi's	kaos	70000	10	63000	L	layak	Cash	3	425
6	12/05/2024	Minggu	2	5	Levi's	jeans	100000	10	90000	L	Bagus	Transfer Bank	1	425
7	12/05/2024	Minggu	2	5	Adidas	kaos	140000	0	140000	M	Seperti Baru	Cash	2	425
8	13/05/2024	Senin	2	5	Uniqlo	Kemeja	100000	10	90000	M	layak	Cash	2	425
9	13/05/2024	Senin	2	5	H&M	Kemeja	140000	10	126000	L	Bagus	Transfer Bank	2	425
10	13/05/2024	Senin	2	5	Gildan	kaos	150000	10	135000	M	Bagus	E-Wallet	3	425
11	13/05/2024	Senin	2	5	Uniqlo	kaos	100000	0	100000	M	layak	Cash	2	425
12	13/05/2024	Senin	2	5	H&M	Baju Polo	120000	10	108000	L	layak	E-Wallet	3	425
13	13/05/2024	Senin	2	5	Levi's	Jacket	120000	10	108000	XL	Seperti Baru	Cash	1	425
14	13/05/2024	Senin	2	5	Uniqlo	Celana Pendek	150000	0	150000	M	layak	E-Wallet	2	425
15	13/05/2024	Senin	2	5	H&M	Hoodie	120000	0	120000	M	layak	E-Wallet	3	425
16	13/05/2024	Senin	2	5	Lacoste	Jacket	100000	0	100000	XL	layak	E-Wallet	1	425
17	13/05/2024	Senin	2	5	Adidas	Jacket	120000	10	108000	XL	Bagus	Transfer Bank	1	425
18	13/05/2024	Senin	2	5	Adidas	Baju Polo	80000	10	72000	L	Bagus	Cash	2	425
19	14/05/2024	Selasa	2	5	Nike	Jacket	100000	10	90000	M	Seperti Baru	Cash	1	425
20	14/05/2024	Selasa	2	5	Dickies	Kemeja	140000	0	140000	XL	Seperti Baru	Transfer Bank	2	425
21	14/05/2024	Selasa	2	5	Dickies	Jeans	70000	10	63000	S	Seperti Baru	Transfer Bank	1	425
22	14/05/2024	Selasa	2	5	Uniqlo	Hoodie	80000	0	80000	S	Seperti Baru	Cash	3	425
23	14/05/2024	Selasa	2	5	Levi's	kaos	110000	10	99000	L	Bagus	Cash	1	425
24	14/05/2024	Selasa	2	5	Dickies	Baju Polo	90000	10	81000	M	Seperti Baru	E-Wallet	2	425

Gambar 4.1 Hasil Pengumpulan Data

4.2. Import Library

```
Import Library
pandas sebagai pd
numpy sebagai np
LabelEncoder dari sklearn.preprocessing
mean_squared_error, mean_absolute_error dari sklearn.metrics
GridSearchCV dari sklearn.model_selection
XGBRegressor dari xgboost
```

Potongan kode di atas terlihat beberapa library penting yang digunakan untuk membangun model *machine learning*. Pandas digunakan untuk membaca, mengelola, serta menganalisis data dalam bentuk tabel agar lebih mudah diproses. NumPy berfungsi dalam melakukan operasi numerik serta manipulasi array dengan efisien. LabelEncoder dari sklearn.preprocessing dimanfaatkan untuk mengubah data kategori berbentuk teks menjadi representasi numerik sehingga dapat dipahami oleh algoritma *machine learning*. Kemudian, dari sklearn.metrics terdapat mean_squared_error (MSE) dan mean_absolute_error (MAE) yang berfungsi sebagai metrik evaluasi untuk mengukur tingkat kesalahan prediksi pada model regresi. Selain itu, GridSearchCV dari sklearn.model_selection digunakan untuk melakukan optimasi parameter model dengan teknik pencarian grid, sehingga dapat menemukan kombinasi parameter terbaik. Terakhir, XGBRegressor dari library XGBoost merupakan algoritma berbasis gradient boosting yang sangat efektif dalam membangun model regresi dengan performa tinggi dan akurasi yang optimal.

4.3. Import Dataset

Load Dataset
Baca file Excel "DATA PENJUALAN CCHASE.xlsx"
Simpan ke dalam variabel dataframe (df)

Pada tahap Load Data digunakan perintah pd.read_excel ('DATA PENJUALAN CCHASE.xlsx') dari *library* pandas untuk membaca file Excel dengan nama DATA PENJUALAN CCHASE.xlsx dan menyimpannya ke dalam variabel df dalam bentuk Dataframe, sehingga data tersebut dapat dengan mudah diolah, dianalisis, maupun diproses lebih lanjut oleh algoritma *machine learning*.

4.4. Pre-Procesing Data

```
 Encoding semua fitur kategorikal ===
 le_kategori = LabelEncoder()
df['Kategori_encoded'] = le_kategori.fit_transform(df['Kategori'])

le_merk = LabelEncoder()
df['Merk_encoded'] = le_merk.fit_transform(df['Merk'])

le_hari = LabelEncoder()
df['Hari_encoded'] = le_hari.fit_transform(df['Hari'])

le_metode = LabelEncoder()
df['Metode_encoded'] = le_metode.fit_transform(df['Metode Pembayaran'])

le_size = LabelEncoder()
df['Size_encoded'] = le_size.fit_transform(df['Size'])

le_kondisi = LabelEncoder()
df['Kondisi_encoded'] = le_kondisi.fit_transform(df['Kondisi'])

Fitur & Target
X_cols = [
   'Minggu Ke','Kategori_encoded','Harga',
   'Harga Akhir','Diskon (%)','Metode_encoded'
]
y_col = 'Total Penjualan'
```

Encoding Fitur Kategorikal

Buat objek LabelEncoder untuk setiap kolom kategorikal

Encode kolom "Kategori" → "Kategori_encoded"

Encode kolom "Merk" → "Merk_encoded"

Encode kolom "Hari" → "Hari_encoded"

Encode kolom "Metode Pembayaran" → "Metode_encoded"

Encode kolom "Size" → "Size_encoded"

Encode kolom "Kondisi" → "Kondisi_encoded"

Tentukan Fitur dan Target

Tentukan variabel input (X_cols) yang digunakan untuk melatih model:

Minggu Ke

Kategori_encoded

Harga

Harga Akhir

Diskon (%)

Metode_encoded

Tentukan variabel output (y_col) sebagai target prediksi:

Total Penjualan.

4.5. Data Preparation

```
Train-Test Split (minggu terakhir = test) ===
train = df[df['Minggu Ke'] < df['Minggu Ke'].max()]
test = df[df['Minggu Ke'] == df['Minggu Ke'].max()]

X_train = train[X_cols]
y_train = train[y_col]

X_test = test[X_cols]
y_test = test[y_col]
```

Split Data Training dan Testing

Training = data dengan "Minggu Ke" < minggu terakhir

Testing = data dengan "Minggu Ke" = minggu terakhir

X_train = data train pada kolom X_cols

y _train = data train pada kolom y_col

X_test = data test pada kolom X_cols

y_test = data test pada kolom y_col

Tahap selanjutnya adalah *train-test* split yang termasuk dalam *data preparation*. Pada kode tersebut, pembagian data dilakukan dengan pendekatan *time-based* split, yaitu menggunakan urutan waktu (Minggu Ke) sebagai acuan, di mana seluruh minggu kecuali minggu terakhir digunakan sebagai data latih (*train*), sedangkan minggu terakhir dijadikan data uji (test). Pendekatan ini dipilih karena data bersifat runtun waktu (*time series*), sehingga lebih tepat dibanding pembagian data secara acak dengan rasio tertentu (misalnya 80:20) yang umumnya digunakan pada data non-time series. Selanjutnya, variabel fitur (X) dan target (y) juga dipisahkan, sehingga diperoleh X_train dan y_train untuk melatih model, serta X_test dan y_test untuk menguji performa model terhadap data baru yang belum pernah dilihat sebelumnya.

4.6 Modelling XGBoost

Model Training dengan Grid Search dan Cross Validation

Definisikan model XGBRegressor dengan objective = "reg:squarederror" dan random_state = 42

Definisikan parameter grid:

n_estimators = [100, 200, 400]
learning_rate = [0.01, 0.05, 0.1]
max_depth = [3, 5, 6, 8]
subsample = [0.8, 1.0]
colsample_bytree = [0.8, 1.0]

Buat objek GridSearchCV dengan:

estimator = xgb_model
param_grid = param_grid
scoring = neg_mean_squared_error
cv = 3
verbose = 1
n_jobs = -1

Fit GridSearchCV ke data training (X_train, y_train)

Simpan best_estimator_ sebagai model terbaik

Cetak parameter terbaik (best_params_)

hyperparameter yang diuji antara lain n_estimators yang menentukan jumlah pohon, learning_rate yang mengatur kecepatan pembelajaran, max_depth yang mengontrol kedalaman pohon, serta subsample dan colsample_bytree yang berfungsi mengatur proporsi data maupun fitur yang digunakan dalam pelatihan. Proses pencarian parameter dilakukan dengan validasi silang (cv=3) agar hasil model lebih optimal dan tidak overfitting. Setelah seluruh kombinasi diuji, GridSearchCV memilih model terbaik berdasarkan nilai error terkecil, lalu model tersebut dilatih menggunakan data latih X_train dan y_train untuk mempelajari hubungan antara fitur input dengan target Total Penjualan.

4.7 Evaluasi Model

Evaluasi (RMSE & MAE)
y_pred_test = model.predict(X_test)
rmse = np.sqrt(mean_squared_error(y_test, y_pred_test))
mae = mean_absolute_error(y_test, y_pred_test)

Gunakan model terbaik untuk memprediksi data uji (X_test).

Simpan hasil prediksi ke variabel y_pred_test.

Hitung nilai Root Mean Squared Error (RMSE):

Kurangi nilai aktual (y_test) dengan hasil prediksi (y_pred_test).

Kuadratkan selisih setiap data.

Hitung rata-rata dari semua selisih kuadrat.

Ambil akar kuadrat dari hasil tersebut.

Simpan hasilnya ke variabel rmse.

Hitung nilai Mean Absolute Error (MAE):

Kurangi nilai aktual (y_test) dengan hasil prediksi (y_pred_test).

Ambil nilai absolut dari setiap selisih.

Hitung rata-rata dari semua selisih absolut.

Simpan hasilnya ke variabel mae.

4.8 Hasil Prediksi

```
Prediksi 1 bulan ke depan (4 minggu)
test_next = test.copy()
test_next['Minggu Ke'] = test_next['Minggu Ke'] + 4  # loncat 4 minggu

X_next = test_next[X_cols]
y_pred = model.predict(X_next)

# tambahkan hasil prediksi
test_next['Prediksi Penjualan'] = np.round(y_pred).astype(int)
test_next['Kategori'] = le_kategori.inverse_transform(test_next['Kategori_encoded'])

# agregasi ke level kategori untuk bulan depan (SEMUA kategori ditampilkan)
result = (
    test_next.groupby('Kategori')['Prediksi  Penjualan']
    .sum()
    .reset_index()
    .sort_values(by='Prediksi Penjualan', ascending=False)
    .reset_index(drop=True)
)

print("\n HASIL EVALUASI ")
print(f"RMSE: {rmse:.4f}")
print(f"MAE: {mae:.4f}\n")

print("Prediksi Penjualan Semua Kategori Bulan Depan ")
print("Kategori      Prediksi Penjualan")
for i, row in result.iterrows():
   print(f"{i+1}. {row['Kategori']:<10} {row['Prediksi Penjualan']}")
```

Buat salinan data uji (test) dan simpan sebagai test_next.

Update kolom "Minggu Ke":

Tambahkan nilai 4 pada setiap baris (loncat 4 minggu ke depan).

Siapkan data input untuk prediksi:

```
X_next = test_next[X_cols]
```

Lakukan prediksi penjualan dengan model terbaik:

```
y_pred = model.predict(X_next)
```

Tambahkan hasil prediksi ke test_next:

Kolom "Prediksi Penjualan" = pembulatan dari y_pred ke bilangan bulat.

Kolom "Kategori" = ubah kembali kode kategori (Kategori_encoded) ke label asli dengan inverse transform.

Agregasi hasil prediksi ke level kategori:

Kelompokkan test_next berdasarkan kolom "Kategori".

Hitung jumlah total "Prediksi Penjualan" tiap kategori.

Reset index hasil agregasi.

Urutkan hasil berdasarkan "Prediksi Penjualan" dari yang terbesar ke terkecil.

Simpan hasil ke variabel result.

Tampilkan hasil evaluasi:

Cetak nilai RMSE.

Cetak nilai MAE.

Tampilkan hasil prediksi penjualan semua kategori bulan depan:

Cetak daftar kategori beserta total prediksi penjualannya.

Urutkan dari prediksi tertinggi ke terendah.

Terakhir, hasil evaluasi model berupa RMSE dan MAE ditampilkan, diikuti dengan daftar prediksi penjualan semua kategori pada bulan depan secara berurutan.

HASIL EVALUASI
RMSE: 0.8105
MAE: 0.6643

Prediksi Penjualan Semua Kategori Bulan Depan
Kategori Prediksi Penjualan
1. Crewneck 118
2. Hoodie 116
3. kaos 114
4. Jacket 100
5. Jeans 98
6. Kemeja 74
7. Baju Polo 70
8. Celana Pendek 8

Dari hasil di atas terlihat bahwa model memiliki performa evaluasi yang cukup baik, dengan nilai RMSE sebesar 0.8105 dan MAE sebesar 0.6643. Nilai tersebut menunjukkan bahwa tingkat kesalahan prediksi relatif kecil sehingga model dapat diandalkan untuk memproyeksikan penjualan. Selanjutnya, dilakukan prediksi penjualan untuk periode 1 bulan ke depan (4 minggu) dengan hasil agregasi ke setiap kategori produk. Hasil prediksi penjualan seluruh kategori adalah sebagai berikut:

a) Crewneck dengan prediksi penjualan sebanyak 118 item,
b) Hoodie dengan prediksi penjualan sebanyak 116 item,
c) Kaos dengan prediksi penjualan sebanyak 114 item,
d) Jacket dengan prediksi penjualan sebanyak 100 item,
e) Jeans dengan prediksi penjualan sebanyak 98 item,
f) Kemeja dengan prediksi penjualan sebanyak 74 item,
g) Baju Polo dengan prediksi penjualan sebanyak 70 item,
h) Celana Pendek dengan prediksi penjualan sebanyak 8 item.

Dengan demikian, kategori Crewneck, Hoodie, dan Kaos menempati posisi teratas, sementara kategori Celana Pendek diproyeksikan memiliki penjualan paling rendah di bulan mendatang.

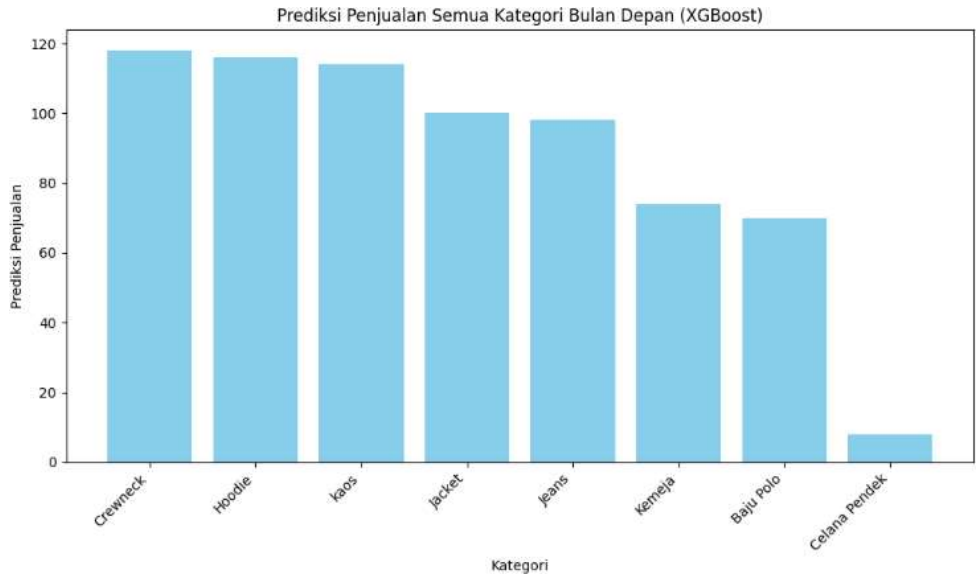
4.9 Visualisasi Prediksi

```
Visualisasi Prediksi Penjualan Semua Kategori Bulan Depan ===
plt.figure(figsize=(10,6))
plt.bar(result['Kategori'], result['Prediksi Penjualan'], color='skyblue')
plt.xticks(rotation=45, ha='right')
plt.xlabel("Kategori")
plt.ylabel("Prediksi Penjualan")
plt.title("Prediksi Penjualan Semua Kategori Bulan Depan (XGBoost)")
plt.tight_layout()
plt.show()

plt.figure(figsize=(10,6))
plt.plot(np.cumsum(y_test.values), label="Aktual (Kumulatif)", marker="o")
plt.plot(np.cumsum(y_pred_test), label="Prediksi (Kumulatif)", marker="s")
plt.xlabel("Index Data Test")
plt.ylabel("Total Kumulatif Penjualan")
plt.title("Perbandingan Kumulatif Aktual vs Prediksi")
plt.legend()
plt.show()
```

Kode diatas menampilkan dua jenis visualisasi hasil prediksi penjualan dengan library Matplotlib. Visualisasi pertama menggunakan diagram batang (bar chart) untuk menunjukkan prediksi penjualan per kategori produk pada bulan depan. Ukuran grafik diatur dengan plt.figure(figsize=(10,6)), sedangkan fungsi plt.bar digunakan untuk membuat batang berwarna skyblue. Label sumbu X adalah “Kategori” dan sumbu Y adalah “Prediksi Penjualan”, dengan judul grafik “Prediksi Penjualan Semua Kategori Bulan Depan (XGBoost)”. Agar label kategori pada sumbu X lebih mudah dibaca, digunakan plt.xticks(rotation=45, ha='right'). Visualisasi kedua menampilkan perbandingan antara data aktual dan prediksi penjualan dalam bentuk kumulatif menggunakan grafik garis (line chart). Fungsi

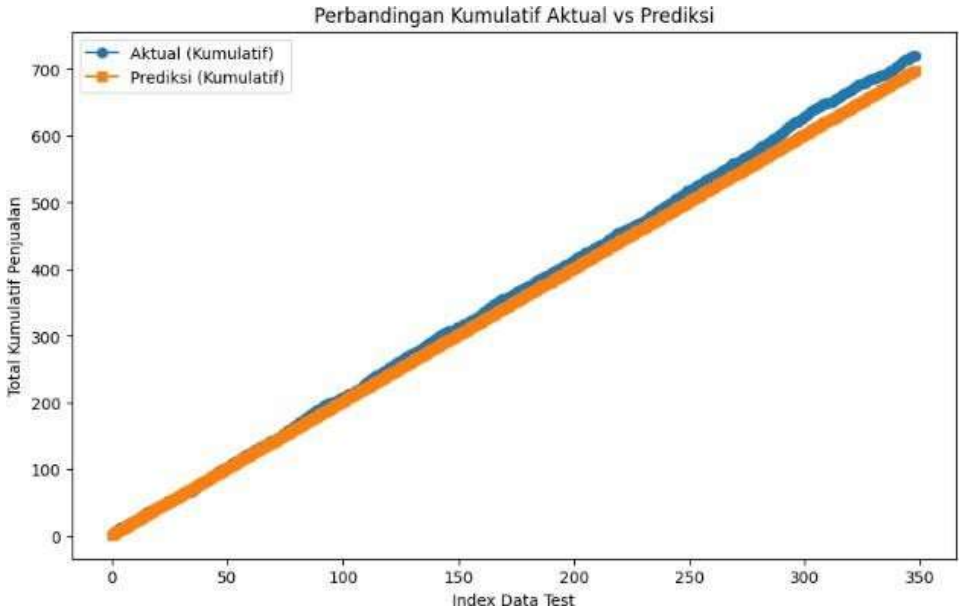
np.cumsum() digunakan untuk menjumlahkan data secara akumulatif sehingga pola tren penjualan lebih terlihat. Garis pertama mewakili data aktual dengan marker lingkaran, sedangkan garis kedua menampilkan prediksi dengan marker kotak. Judul, label sumbu, serta legenda ditambahkan agar grafik mudah dipahami. Dengan dua visualisasi ini, pembaca dapat memperoleh gambaran jelas baik mengenai distribusi prediksi per kategori maupun kesesuaian tren prediksi terhadap data aktual.



Gambar 4.2 Bar Chart Prediksi Penjualan

Gambar 4.3 menampilkan hasil visualisasi prediksi penjualan untuk semua kategori produk pada bulan depan menggunakan algoritma XGBoost. Visualisasi ditampilkan dalam bentuk diagram batang (*bar chart*) dengan warna biru muda (*skyblue*) untuk memberikan tampilan yang lebih jelas. Sumbu X merepresentasikan kategori produk seperti Crewneck, Hoodie, Kaos, Jacket, Jeans, Kemeja, Baju Polo, dan Celana Pendek, sedangkan sumbu Y menunjukkan jumlah prediksi penjualan pada masing-masing kategori. Dari grafik ini terlihat bahwa kategori dengan penjualan tertinggi adalah Crewneck, Hoodie, dan Kaos yang

diprediksi terjual lebih dari 110 item, sedangkan kategori dengan prediksi penjualan terendah adalah Celana Pendek dengan jumlah kurang dari 10 item. Visualisasi ini memberikan gambaran yang lebih mudah dipahami mengenai perbandingan prediksi penjualan antar kategori produk, sehingga dapat digunakan sebagai acuan dalam pengambilan keputusan strategi penjualan dan pengelolaan stok barang.



Gambar 4.3 Line Chart Perbandingan Nilai Aktual dan Prediksi

Dari gambar di atas terlihat visualisasi perbandingan kumulatif nilai aktual dan prediksi penjualan. Garis biru merepresentasikan data aktual kumulatif, sedangkan garis oranye menunjukkan hasil prediksi kumulatif dari model XGBoost. Pola kedua garis terlihat sangat berdekatan dan hampir sejajar dari awal hingga akhir, yang menandakan bahwa model mampu mengikuti tren penjualan dengan baik. Perbedaan yang muncul antara garis aktual dan prediksi relatif kecil, sehingga dapat disimpulkan bahwa model memiliki performa prediksi yang cukup akurat dalam menggambarkan pola kumulatif penjualan.

BAB V

PENUTUP

5.1 Kesimpulan

Dari penelitian yang telah dilakukan dapat disimpulkan:

1. Model *machine learning* dengan algoritma XGBoost berhasil dibangun untuk melakukan prediksi penjualan berdasarkan data historis yang tersedia.
2. Hasil evaluasi menunjukkan nilai RMSE sebesar 0.8105 dan MAE sebesar 0.6643, yang berarti model mampu melakukan prediksi dengan tingkat kesalahan yang relatif kecil sehingga cukup andal digunakan.
3. Berdasarkan hasil prediksi penjualan untuk periode 1 bulan ke depan (4 minggu), diperoleh tiga kategori dengan prediksi penjualan tertinggi, yaitu:
   a) Crewneck dengan 118 item
   b) Hoodie dengan 116 item
   c) Kaos dengan 114 item

   Hal ini menunjukkan bahwa ketiga kategori tersebut diperkirakan menjadi penyumbang terbesar terhadap total penjualan di bulan berikutnya.

4. Penelitian ini terletak pada penerapan algoritma XGBoost pada toko thrifting lokal (Cchase), yang berbeda dari penelitian serupa pada retail besar. Hal ini membuktikan bahwa XGBoost tidak hanya relevan untuk industri berskala besar, tetapi juga dapat diadaptasi secara efektif pada usaha kecil dengan karakteristik data yang unik, seperti variasi stok yang terbatas dan tidak konsisten.

5.2 Saran

Adapun saran yang dapat diberikan oleh penulis untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Penelitian ini masih berfokus pada prediksi penjualan dengan variabel terbatas, sehingga pada penelitian berikutnya disarankan untuk menambahkan variabel lain seperti tren musiman, promosi, maupun faktor eksternal lain agar hasil prediksi lebih akurat.
2. Hasil prediksi yang sudah diperoleh dapat dimanfaatkan sebagai dasar dalam perencanaan strategi bisnis, khususnya untuk pengelolaan stok barang dan penentuan strategi promosi agar lebih tepat sasaran.
3. Membandingkan dengan algoritma lain seperti Random Forest, KNN, Decision Tree, SVM, maupun metode berbasis deep learning (misalnya CNN atau LSTM untuk data deret waktu) sebagai baseline perbandingan performa.

Dengan adanya saran-saran ini, diharapkan penelitian di bidang prediksi penjualan berbasis *machine learning* dapat terus berkembang dan memberikan kontribusi yang lebih besar dalam dunia bisnis, khususnya dalam membantu perusahaan menyusun strategi penjualan, manajemen stok, serta promosi yang lebih efektif berdasarkan analisis data historis.

DAFTAR PUSTAKA

Adji, N. L., & Claretta, D. (2022). Fenomena Thrift Shop Dikalangan Remaja:Studi Fenomenologi tentang Thrift Shop di Kalangan Remaja Surabaya. *Journal Of Communication And Islamic Broadcasting, 2*(4), 36-44.https://doi.org/10.47467/dawatuna.v2i4.2201

Bengnga, A., & Ishak, R. (2023). Penerapan XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah. *Jambura Journal of Electrical and Electronics Engineering, 5*(2), 192–196. https://doi.org/10.37905/jjeee.v5i2.20253

Anggara, J., Ramadhan, F. R., Syofian, A. F., Panjaitan, A. A., & Fithra, R. F. (2025). Pengembangan Sistem Prediksi Harga dan Rekomendasi Mobil Bekas Berbasis *Machine learning*. *Journal of Technology and Informatics (JoTI), 7*(1), 11-23 https://doi.org/10.37802/joti.v7i1.987

Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python sebagai bahasa pemrograman untuk *machine learning* dan deep learning. *Karimah Tauhid, 2*(1), 1-6. https://doi.org/10.30997/karimahtauhid.v2i1.7518

Hidayat, M. T., & Sulistiyono, M. (2025). Analisis Performa Algoritma XGBoost, GRU, dan Prophet dalam Peramalan Penjualan Obat untuk Optimasi Rantai Pasok Farmasi. *Jurnal Pendidikan dan Teknologi Indonesia (JPTI), 5*(1), 65–73. https://doi.org/10.52436/1.jpti.562

Jury, C. W. A., & Fimelia, Q. (2024). Analisis Pengambilan Keputusan Kementerian Perdagangan Republik Indonesia Dalam Larangan Impor Pakaian Bekas (Thrifting) Menurut Teori Rasional. *JURNAL ILMU HUKUM DAN TATA NEGARA, 2*(2), 200-217. https://doi.org/10.55606/birokrasi.v2i2.1185

Kirana, D. H., & Febriola, B. (2025). STRATEGI PENANGGULANGAN DAMPAK LARANGAN PENJUALAN PAKAIAN BEKAS IMPOR DI INDONESIA TERHADAP PEMASUKAN PEDAGANG. *Indonesian Journal of Business and Management, 5*(1), 58-68. https://doi.org/10.53363/buss.v5i1.337

Nadhila, S., Muzhirah, M., Sajali, H., & Andinata, M. (2023). Eksistensi Diri Remaja Dalam Penggunaan Pakaian Bekas (Studi Kasus Pada Konsumen Thrifting Pajak Melati Medan). *Journal Of Social Science Research, 3*(3), 2436-2446. https://j-innovative.org/index.php/Innovative

Nurak, Y. M., Irianandа, S. W., & Marisa, F. (2025). PREDIKSI PENJUALAN WARUNG KOPI OI MENGGUNAKAN METODE RANDOM FOREST DAN XGBOOST. *JATI (Jurnal Mahasiswa Teknik Informatika), 9*(4), 5796-5802. https://doi.org/10.36040/jati.v9i4.13922

Nurmaya, S. T., Soesilo, B., & Fatah, D. A. (2024). Prediksi harga saham pada Bank BCA Tbk dengan teknik XGBoost (Extreme Gradient Boosting). *JURSIMA, 12*(2), 370–379. https://doi.org/10.47024/js.v12i2.925

Pratama, A. F., Suryani, D., Izzudin, F. S., & Adzikirani, A. (2024). ANALISIS PREDIKTIF MENGGUNAKAN *MACHINE LEARNING* UNTUK MENANGGULANGI MASALAH REJECT PRODUK PADA PROSES PRODUKSI PT. XYZ. *Journal of Innovation Research and Knowledge, 4*(5), 2971-2984. https://bajangjournal.com/index.php/JIRK/article/view/8725

Riza, F. (2022). Analisis dan prediksi data penjualan menggunakan *machine learning* dengan pendekatan ilmu data. *Data Sciences Indonesia, 1*(2), 62-68. https://doi.org/10.47709/dsi.v1i2.1308

Rayadin, M. A., Musaruddin, M., Saputra, R. A., & Isnawaty, I. (2024). Implementasi Ensemble Learning Metode XGBoost dan Random Forest untuk Prediksi Waktu Penggantian Baterai Aki. *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer, 5*(2), 111-119. https://doi.org/10.37148/bios.v5i2.128

Santika, D. N. A. W., & Bagiastra, I. N. (2021). LEGALITAS THRIFT SHOP DAN PRELOVED DI INDONESIA. *Jurnal Kertha Desa*, *9*(6), 24-33. https://ojs.unud.ac.id/index.php/kerthadesa/article/view/70205

Siringoringo, R., Perangin-Angin, R., & Purba, M. J. (2021). Segmentasi Dan Peramalan Pasar Retail Menggunakan XGBoost Dan Principal Component Analysis. *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt*, *5*(1), 42-47. https://doi.org/10.46880/jmika.Vol5No1.pp42-47

Saputro, M. R., Mahdiyah, U., & Swanjaya, D. (2024). Perbandingan Metode Adaptive Boosting dan Extreme Gradient Boosting Untuk Prediksi Hasil Pertandingan Liga Spanyol. *Nusantara of Engineering (NOE)*, *7*(1), 67-73. https://doi.org/10.29407/noe.v7i1.20882

Tiastama, S. A., & Budi, I. (2024). Perbandingan random search dan algoritma genetika dalam penyetelan hyperparameter XGBoost pada retail sales forecasting. *The Indonesian Journal of Computer Science*, *13*(4), 6602–6613. https://doi.org/10.33022/ijcs.v13i4.4285

Utarsih, H. (2024). Analisis Kualitas Produk dan Harga Terhadap Keputusan Pembelian pada Thrifting Pasar Cimol Gedebage. *Journal of Innovation inManagement, Accounting and Business, 3*(2), 170-177. https://doi.org/10.56916/jimab.v3i2.909

Widiana, S. A., Firdaus, I., Tenda, E., & Ketaren, E. (2023). Sistem Informasi Prediksi Penjualan Produk Thrift Di Toko Manado Menggunakan Algoritma K-Nearest Neighbor (K-Nn). *Jurnal TIMES, 12*(2), 52-57. https://doi.org/10.51351/jtm.12.2.2023708

Wijaya, H., Hostiadi, D. P., & Triandini, E. (2024). Meningkatkan prediksi penjualan retail XYZ dengan teknik optimasi random search pada model XGBoost. *Jurnal Teknologi dan Informasi STIKOM Bali, 1*(2), 829–833. https://spinter.stikom-bali.ac.id/index.php/spinter/article/view/226

Winurputra, R., & Ratnawati, D. E. (2025). Peramalan penjualan produk menggunakan Extreme Gradient Boosting (XGBoost) dan kerangka kerja CRISP-DM untuk pengoptimalan manajemen persediaan (studi kasus: UB Mart). *Jurnal Teknologi Informasi dan Ilmu Komputer, 12*(2), 417–428. https://doi.org/10.25126/jtiik.2025129451

Yulianti, S. E. H., Soesanto, O., & Sukmawaty, Y. (2022). Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit. *Journal of Mathematics: Theory and Applications, 4*(1), 21-26. https://doi.org/10.31605/jomta.v4i1.1792

SKRIPSI_MUHAMMAD_ICHWAN_GIFARI_DONE-1758458080212

ORIGINALITY REPORT

23%	19%	12%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	repository.umsu.ac.id Internet Source	2%
2	Submitted to Universitas Muhammadiyah Sumatera Utara Student Paper	1%
3	digilib.uin-suka.ac.id Internet Source	1%
4	ojs.unud.ac.id Internet Source	1%
5	Amiruddin Bengnga, Rezqiwati Ishak. "Penerapan XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah", Jambura Journal of Electrical and Electronics Engineering, 2023 Publication	<1%
6	Vernanda Sam Saputra, Achmad Ridwan, Taftazani Ghazi Pratama. "RANCANG BANGUN SISTEM REKOMENDASI BUKU BERBASIS ITEM-BASED COLLABORATIVE FILTERING MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS", Jurnal Informatika dan Teknik Elektro Terapan, 2025 Publication	<1%
7	j-innovative.org Internet Source	<1%

8	Submitted to Universitas Sumatera Utara Student Paper	< 1 %
9	journal-stiayappimakassar.ac.id Internet Source	< 1 %
10	Yanuarini Nur Sukmaningtyas, Ronny Makhfuddin Akbar, Gita Rohma Utami Asyafiiyah. "Penerapan Predictive Analytics untuk Analisis Faktor-faktor yang Mempengaruhi Performa Akademik Siswa", Arcitech: Journal of Computer Science and Artificial Intelligence, 2024 Publication	< 1 %
11	doaj.org Internet Source	< 1 %
12	core.ac.uk Internet Source	< 1 %
13	Dwi Nurmelly Handayani, Sayid Qutub. "Penerapan Random Forest Untuk Prediksi Dan Analisis Kemiskinan", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025 Publication	< 1 %
14	jurnal.itscience.org Internet Source	< 1 %
15	e-journals.dinamika.ac.id Internet Source	< 1 %
16	ojs.unpkediri.ac.id Internet Source	< 1 %
17	Ridho Ananda Purba, Husni Ritonga. "Pemanfaatan Media Sosial Sebagai Wadah Usaha Pakaian Bekas di Pajak Melati Kota	< 1 %

	Medan: Studi Kasus pada Aplikasi Tiktok", Reslaj: Religion Education Social Laa Roiba Journal, 2024 Publication	
18	Rizal Muslim Sinaga, Muhammad Rosyid Fauzan, Ega Pratama, Muhammad Rizki Alfahri, Kana Saputra. "Pemanfaatan Teknik Machine Learning dalam Memprediksi Kepadatan Lalu Lintas Guna Efisiensi Transportasi", PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer, 2025 Publication	<1%
19	Sheren Afryan Tiastama, Indra Budi. "Perbandingan Random Search dan Algoritma Genetika dalam Penyetelan Hyperparameter XGBoost pada Retail Sales Forecasting", The Indonesian Journal of Computer Science, 2024 Publication	<1%
20	medium.com Internet Source	<1%
21	ioinformatic.org Internet Source	<1%
22	repositori.umrah.ac.id Internet Source	<1%
23	repository.ubharajaya.ac.id Internet Source	<1%
24	spinter.stikom-bali.ac.id Internet Source	<1%
25	www.researchsquare.com Internet Source	<1%

26	Submitted to UM Surabaya Student Paper	<1%
27	bios.sinergis.org Internet Source	<1%
28	Submitted to University of Liverpool Student Paper	<1%
29	artikelpendidikan.id Internet Source	<1%
30	ejournal.stmik-time.ac.id Internet Source	<1%
31	repo.darmajaya.ac.id Internet Source	<1%
32	repository.uin-suska.ac.id Internet Source	<1%
33	etheses.uin-malang.ac.id Internet Source	<1%
34	www.grafiati.com Internet Source	<1%
35	Submitted to Universitas Mohammad Husni Thamrin Jakarta Student Paper	<1%
36	elibs.unigres.ac.id Internet Source	<1%
37	eprints.amikompurwokerto.ac.id Internet Source	<1%
38	repositori.usu.ac.id Internet Source	<1%
39	rodin.uca.es Internet Source	<1%

40	Submitted to Coventry University Student Paper	<1%
41	mathjournal.unram.ac.id Internet Source	<1%
42	ojs.unida.ac.id Internet Source	<1%
43	repository.unjaya.ac.id Internet Source	<1%
44	repository.unuja.ac.id Internet Source	<1%
45	ojs.trigunadharma.ac.id Internet Source	<1%
46	Submitted to Tarumanagara University Student Paper	<1%
47	jiped.org Internet Source	<1%
48	Neternomia Maria Ausiladora Da Costa, Anief Fauzan Rozi. "Analisis efisiensi tidur berdasarkan faktor demografi dan kebiasaan harian dengan metode random forest regression", Jurnal Pendidikan Informatika dan Sains, 2025 Publication	<1%
49	Submitted to Sekolah Teknik Elektro & Informatika Student Paper	<1%
50	Susliansyah Susliansyah, Sigit Yugi Wargiyo, Heny Sumarno, Hendro Priyono, Linda Maulida. "Rancangan Aplikasi Algoritma C4.5	<1%

	pada Stunting Balita Menggunakan Bahasa Phyton", remik, 2025 Publication	
51	T. Mariprasath, Kumar Reddy Cheepati, Marco Rivera. "Practical Guide to Machine Learning, NLP, and Generative AI: Libraries, Algorithms, and Applications", River Publishers, 2024 Publication	< 1 %
52	ojs.unsulbar.ac.id Internet Source	< 1 %
53	onlinetutorialspoint.com Internet Source	< 1 %
54	scholar.ummetro.ac.id Internet Source	< 1 %
55	www.scribd.com Internet Source	< 1 %
56	Andre Wiejaya, Indah Fenriana. "Prediksi Harga Saham Top 10 NASDAQ dengan Time Series Prophet", bit-Tech, 2024 Publication	< 1 %
57	Submitted to RMIT University Student Paper	< 1 %
58	Submitted to University of Wolverhampton Student Paper	< 1 %
59	eprints.uns.ac.id Internet Source	< 1 %
60	jurnal.stkippgritulungagung.ac.id Internet Source	< 1 %
61	repository.untag-sby.ac.id Internet Source	< 1 %

62	www.mdpi.com Internet Source	<1%
63	Submitted to Universitas Bengkulu Student Paper	<1%
64	Submitted to Universitas Esa Unggul Student Paper	<1%
65	Submitted to Universitas Muslim Indonesia Student Paper	<1%
66	digilib.unila.ac.id Internet Source	<1%
67	docplayer.info Internet Source	<1%
68	ejournal.raharja.ac.id Internet Source	<1%
69	www.slideshare.net Internet Source	<1%
70	yganalyst.github.io Internet Source	<1%
71	etd.repository.ugm.ac.id Internet Source	<1%
72	journal.politeknik-pratama.ac.id Internet Source	<1%
73	repository.its.ac.id Internet Source	<1%
74	Lely Kurniawati, Dadang Priyanto, Neny Sulistia Ningsih, Moch Syahrir, Ria Rismayati. "Comparison of Decision Tree-Based Methods in Lung Disease Detection", Jurnal Bumigora Information Technology (BITe), 2025 Publication	<1%

75	Muhammad Wisnu Nugroho. "Analisis Performa Algoritma Random Forest dalam Mengatasi Overfitting pada Model Prediksi", Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi), 2025 Publication	<1%
76	datascienceomar.wordpress.com Internet Source	<1%
77	etd.umy.ac.id Internet Source	<1%
78	putrisardyoriza.blogspot.com Internet Source	<1%
79	repository.unhas.ac.id Internet Source	<1%
80	vdokumen.com Internet Source	<1%
81	Belqis Destinola Maritza, Didik Hariyanto. "Strategi Pemasaran Digital untuk Bisnis Pakaian Thrift (akun @spezialmbois)", Interaction Communication Studies Journal, 2024 Publication	<1%
82	Submitted to Brunel University Student Paper	<1%
83	Submitted to STT PLN Student Paper	<1%
84	Sebastian Ubaidillah Royan, Nana Suarna, Irfan Ali, Dodi Solihudin. "ANALISIS SENTIMEN ULASAN PRODUK SKINCARE DI SHOPEE UNTUK MENINGKATKAN KUALITAS PRODUK MENGGUNAKAN METODE SUPPORT VECTOR	<1%

MACHINE", Jurnal Informasi dan Komputer, 2025
Publication

85	ejurnal.ung.ac.id Internet Source	<1%
86	ekaferiana.blogspot.com Internet Source	<1%
87	eprints.politeknikpu.ac.id Internet Source	<1%
88	eprints.uad.ac.id Internet Source	<1%
89	fr.scribd.com Internet Source	<1%
90	jpti.journals.id Internet Source	<1%
91	jurnal.syntaxliterate.co.id Internet Source	<1%
92	repository.teknokrat.ac.id Internet Source	<1%
93	repository.ub.ac.id Internet Source	<1%
94	sistemasi.ftik.unisi.ac.id Internet Source	<1%
95	tdctoolkit.org Internet Source	<1%
96	www.theseus.fi Internet Source	<1%
97	yhudiyuba.blogspot.com Internet Source	<1%

98	Usman Qamar, Muhammad Summair Raza. "Applied Text Mining", Springer Science and Business Media LLC, 2024 Publication	<1%
99	www.numberanalytics.com Internet Source	<1%
100	Abdullah Marshal Permana. "Perlindungan Konsumen Terhadap Korban Penipuan Online Jual Beli Pakaian Bekas Impor (Thrifting)", Proceedings Series on Social Sciences & Humanities, 2024 Publication	<1%
101	Hermanda Asmarani, Suyanto Suyanto, Jawoto Nusantoro. "Implementasi Model Machine Learning dalam Membaca Data Foreign Sell pada Saham", Expensive: Jurnal Akuntansi dan Keuangan, 2025 Publication	<1%
102	Muhammad Miftakhudin, Aang Alim Murtopo, Zaenul Arif. "Integrasi Artificial Neural Network dan Algoritma Genetika untuk Prediksi Bencana Banjir Pesisir Kota Tegal", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		