

SISTEM DATABASE

INGGRID THERESIA SITANGGANG | 241712080

DAFFA INDRASYAH DAULAY | 241712077

BAIDHOWI SIAGIAN | 241712064

KOM B'24



**PROGRAM STUDI D3-TEKNIK INFORMATIKA
FAKULTAS VOKASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2025**

KATA PENGANTAR

Puji dan syukur kami panjatkan kepada Allah SWT, dengan limpah karuniaNya kami dapat menyelesaikan penyusunan makalah ini dengan judul Sistem Database. Makalah ini disusun sebagai pemenuhan tugas proyek akhir mata kuliah Praktikum Sistem Basis Data.

Terima kasih kami sampaikan kepada setiap pihak yang sudah mendukung selama berlangsungnya pembuatan makalah ini. Terkhusus lagi kami sampaikan terima kasih kepada Bapak Dr Syahriol Sitorus S.Si., M.IT selaku dosen pengajar dan asisten laboratorium, Cahaya Pratista yang telah membimbing kami selama proses penggerjaan. Semoga makalah ini bisa bermanfaat bagi setiap pembaca.

Disertai keseluruhan rasa rendah hati, kritik dan saran yang membangun amat kami nantikan, agar nantinya kami dapat meningkatkan dan merevisi kembali pembuatan makalah di tugas lainnya dan di waktu berikutnya.

Medan, 3 Desember 2025

Penulis

DAFTAR ISI

	Halaman
KATA PENGANTAR	i
DAFTAR ISI	ii
DAFTAR GAMBAR	iii
DAFTAR TABEL	iv
BAB 1 PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Batasan Masalah.....	2
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penulisan	3
BAB 2 TINJAUAN PUSTAKA	
2.1 Komponen Penelitian	4
2.1.1 E-Commerce.....	5
2.1.2 Database dan Normalisasi	5
2.1.3 <i>Entity-Relationship Diagram (ERD)</i>	6
2.1.4 Kaitan ERD dengan Tujuan Penelitian.....	6
2.1.5 Analisis Data Transaksi	6
2.2 Metode Penggerjaan.....	6
2.2.1 Pengumpulan dan Observasi Dataset	7
2.2.2 Identifikasi Atribut dan Dependensi Fungsional	7
2.2.3 Penerapan Normalisasi(1NF-3NF).....	7
2.2.4 Validasi Struktur Relasional	8
2.2.5 Implementasi ke RDBMS (Opsional/Pengujji Struktur)	8
BAB 3 METODOLOGI PENELITIAN	
3.1 Diagram Penelitian	9
3.1.1 Flowchart.....	10
3.2 Jadwal Penelitian.....	11
BAB 4 HASIL	
4.1 Pengertian	12
4.1.1 Struktur Data Mentah yang Telah Dianalisis.....	12
4.1.2 Hasil Pemecahan Tabel melalui Normalisasi (1NF-3NF)	12
4.1.3 Hasil Penyusunan ERD (<i>Entity-Relationship Diagram</i>)	15
4.1.4 Hasil Akhir Struktur Basis Data Relasional	15
4.2 Hasil Uji Query	16
BAB 5 PENUTUP	
5.1 Kesimpulan.....	24
5.2 Saran.....	24
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 3.1 Flowchart Penelitian.....	10
Gambar 4.1 Hasil Penyusunan ERD (<i>Entity-Relationship Diagram</i>).....	15
Gambar 4.2 Hasil Query pasangan produk yang sering dibeli customer yang sama.	16
Gambar 4.3 Hasil Query produk yang memberi total omzet tertinggi.....	17
Gambar 4.4 Hasil Query Negara yang memberikan total omzet paling besar.....	18
Gambar 4.5 Hasil Query rata rata rating serta jumlah review.....	19
Gambar 4.6 Hasil Query produk yang paling banyak menghasilkan omzet.....	19
Gambar 4.7 Hasil Query Negara yang gendernya paling beragam.....	20
Gambar 4.8 Hasil Query Negara yang memiliki jumlah pelanggan aktif terbanyak.	21
Gambar 4.9 Produk yang paling banyak dibeli tapi paling sedikit direview?	22
Gambar 4.10 Hasil Query tiga produk yang jadi <i>Best Seller?</i>	22
Gambar 4.11 Hasil Query produk yang paling besar variasi harganya.....	23

DAFTAR TABEL

Tabel 3.1 Jadwal Pelaksanaan Penelitian	11
Tabel 4.1 Tabel 1NF	13
Tabel 4.2 Tabel 2NF	14
Tabel 4.3 Tabel 3NF	14

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Di era digital saat ini, data menjadi salah satu sumber daya terpenting, khususnya dalam sistem penjualan berbasis e-commerce. Setiap transaksi yang tercatat menghasilkan informasi bernilai, yang apabila dianalisis dengan tepat dapat memberikan gambaran mengenai perilaku pelanggan dan tren penjualan. Dataset *ecommerce_dataset_10000.csv* yang digunakan dalam proyek ini merupakan contoh nyata dari kumpulan data transaksi yang masih berada dalam bentuk mentah.

Salah satu permasalahan utama dari penyimpanan data mentah dalam satu tabel besar adalah munculnya redundansi data atau pengulangan informasi yang sama. Pada dataset ini, misalnya, atribut pelanggan seperti *gender*, *age group*, dan *country* terus berulang pada setiap transaksi yang dilakukan oleh pelanggan yang sama. Kondisi ini tidak hanya membuat penyimpanan menjadi tidak efisien, tetapi juga dapat menimbulkan anomali data, seperti kesalahan pada saat melakukan pembaruan, penghapusan, atau penambahan data baru (Han, Kamber, & Pei, 2012). Jika ada perubahan data pada satu pelanggan, maka seluruh baris transaksi yang berhubungan harus diperbarui satu per satu, sehingga meningkatkan risiko inkonsistensi.

Untuk menyelesaikan masalah tersebut, diperlukan perancangan basis data yang lebih terstruktur dan efisien. Normalisasi menjadi metode standar yang digunakan untuk memecah satu tabel besar yang redundan menjadi beberapa tabel yang lebih kecil dan saling terhubung secara relasional. Dengan menerapkan normalisasi hingga Bentuk Normal Ketiga (3NF) pada dataset ini, struktur data menjadi lebih rapi, mudah dikelola, serta bebas dari anomali. Proses ini juga menjadi langkah penting sebelum data diimplementasikan ke dalam Sistem Manajemen Basis Data Relasional (RDBMS) seperti MySQL, sehingga data siap digunakan untuk analisis lebih lanjut.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang ditemukan pada dataset e-commerce terkait redundansi data, struktur tabel yang tidak efisien, serta kebutuhan analisis perilaku

pelanggan, maka rumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana struktur data transaksi penjualan pada sistem e-commerce yang terdapat dalam dataset?
2. Bagaimana proses normalisasi yang tepat untuk diterapkan pada dataset transaksi tersebut agar menghasilkan struktur basis data yang efisien?
3. Informasi apa yang dapat dianalisis dari perilaku pelanggg berdasarkan pola transaksi, kategori produk, rating, dan review?

1.3 Batasan Masalah

Agar penelitian ini lebih terfokus dan pembahasannya mendalam, maka penulis menetapkan beberapa batasan masalah sebagai berikut:

1. Dataset yang digunakan hanya berasal dari file “sbd kelompok 3.xlsx”, yang berisi data pelanggan, produk, transaksi, metode pembayaran, rating, dan review. Penelitian tidak menggunakan dataset tambahan di luar file tersebut.
2. Proses normalisasi yang dilakukan dibatasi sampai Bentuk Normal Ketiga (3NF). Normalisasi tingkat lanjut seperti BCNF, 4NF, atau 5NF tidak dibahas dalam penelitian ini.
3. Fokus penelitian hanya pada perancangan struktur basis data, yaitu identifikasi atribut, pemetaan relasi, dan pemecahan tabel berdasarkan aturan normalisasi. Implementasi fisik database di RDBMS seperti MySQL hanya dibahas secara dasar.
4. Analisis data yang dilakukan bersifat deskriptif, terbatas pada pengelompokan kategori produk, metode pembayaran, rating pelanggan, dan informasi umum lainnya. Tidak dilakukan analisis lanjutan seperti machine learning, prediksi penjualan, ataupun analisis sentimen mendalam.
5. Penelitian tidak membahas aspek bisnis atau strategi pemasaran, melainkan hanya fokus pada pengolahan data dan perancangan model database berdasarkan struktur dataset.
6. Validitas data sepenuhnya bergantung pada isi dataset asli, sehingga penelitian tidak melakukan verifikasi akurasi, kelengkapan, atau keabsahan data dari file tersebut.

1.4 Tujuan Penelitian

Selaras dengan rumusan masalah yang telah dijabarkan, tujuan yang ingin dicapai melalui penelitian ini adalah sebagai berikut:

1. Mengidentifikasi dan mendeskripsikan struktur data transaksi penjualan pada dataset e-commerce sebagai dasar untuk memahami atribut, relasi, serta potensi redundansi.
2. Menerapkan proses normalisasi hingga Bentuk Normal Ketiga (3NF) untuk menghasilkan desain basis data yang terstruktur, bebas duplikasi, dan sesuai prinsip database relasional.
3. Menganalisis perilaku pelanggan berdasarkan pola pembelian, kategori produk yang sering dibeli, serta informasi rating dan review agar dapat memberikan insight mengenai preferensi dan pengalaman pelanggan.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat yang praktis, baik bagi kami sebagai penulis maupun bagi pihak lain yang membaca. Manfaat tersebut dirinci sebagai berikut:

1. Sebagai referensi dalam analisis data e-commerce.
2. Memberikan gambaran penerapan normalisasi pada dataset nyata.
3. Menjadi dasar bagi pengembangan sistem informasi penjualan digit

BAB 2

TINJAUAN PUSTAKA

2.1 Komponen Penelitian

Komponen penelitian dalam pengolahan dan normalisasi dataset mencakup beberapa elemen penting yang saling berkaitan. Komponen tersebut terdiri dari perangkat lunak, data mentah, proses pengolahan, serta teknik perancangan basis data yang digunakan untuk menghasilkan struktur database yang efisien. Adapun komponen-komponen penelitian tersebut adalah sebagai berikut:

1. Dataset Transaksi (Data Mentah)

Penelitian ini menggunakan file *sbd kelompok 3.xlsx* sebagai sumber data utama. Dataset ini berisi informasi pelanggan, produk, transaksi penjualan, metode pembayaran, rating, serta review pelanggan. Data mentah ini menjadi landasan untuk proses identifikasi atribut, deteksi redundansi, dan pemecahan tabel.

2. Perangkat Lunak Pengolahan Data

Penelitian memanfaatkan perangkat lunak seperti Microsoft Excel atau Google Sheets untuk eksplorasi awal dataset, pengecekan struktur data, serta identifikasi potensi anomali dan pengulangan. Selain itu, perangkat lunak RDBMS seperti MySQL atau phpMyAdmin digunakan untuk menguji hasil rancangan tabel setelah dinormalisasi (W3Schools, 2024).

3. Metode Normalisasi Basis Data

Normalisasi merupakan komponen inti dalam penelitian ini, yang bertujuan menghilangkan redundansi, meminimalkan anomali data, dan memastikan integritas data. Normalisasi dilakukan secara bertahap mulai dari 1NF, 2NF, hingga 3NF untuk menghasilkan struktur database yang lebih efisien dan relasional.

4. *Entity-Relationship Diagram* (ERD)

ERD digunakan sebagai alat bantu visual untuk memetakan hubungan antar entitas dalam dataset. Melalui ERD, entitas seperti *Customers*, *Products*, *Orders*, *OrderItems*, dan *Reviews* dapat diidentifikasi beserta atribut dan relasinya. Diagram ini menjadi pedoman dalam proses pemecahan tabel dan desain akhir database.

5. Komponen Validasi dan Pengujian Struktur Data

Setelah dilakukan perancangan skema tabel, penelitian mencakup langkah validasi seperti pengecekan dependensi fungsional, konsistensi relasi, dan pengujian query dasar pada RDBMS untuk memastikan bahwa struktur yang dihasilkan sesuai dengan prinsip normalisasi dan bebas anomali.

2.1.1 E-Commerce

E-Commerce (Electronic Commerce) adalah kegiatan jual beli barang atau jasa melalui media elektronik, terutama internet, yang memungkinkan penjual dan pembeli bertransaksi tanpa harus bertemu langsung. Dalam e-commerce, transaksi dilakukan melalui website atau aplikasi dengan fitur seperti katalog produk, keranjang belanja, dan sistem pembayaran digital, baik melalui transfer bank, e-wallet, maupun kartu kredit. Contoh e-commerce meliputi marketplace seperti Tokopedia, Shopee, dan Lazada, serta toko online mandiri dan layanan digital seperti Traveloka. E-commerce memiliki kelebihan berupa jangkauan pasar yang luas, efisiensi waktu dan biaya, serta transaksi yang cepat dan tersedia 24 jam, meskipun juga memiliki kekurangan seperti risiko penipuan, sulitnya melihat barang secara langsung, serta potensi masalah dalam pengiriman. (Laudon & Traver, 2022).

2.1.2 Database dan Normalisasi

Database adalah kumpulan data yang tersusun sehingga dapat diakses dan dikelola secara efisien (Kroenke & Auer, 2019). Normalisasi adalah proses perancangan skema basis data untuk mengurangi redundansi dan meningkatkan integritas data dengan memecah satu tabel besar menjadi beberapa tabel yang saling relasi. Proses normalisasi biasanya dilakukan hingga bentuk-bentuk normal seperti berikut (Connolly & Begg, 2015):

1. 1NF: Menghilangkan multivalue.
2. 2NF: Menghilangkan ketergantungan parsial.
3. 3NF: Menghilangkan ketergantungan transitif.

2.1.3 Entity-Relationship Diagram (ERD)

ERD merupakan alat visual yang digunakan untuk menggambarkan struktur logis basis data-entitas (tabel), atribut (kolom), dan hubungan antar entitas (relasi). ERD membantu perancang basis data memahami dependensi antar data sehingga dapat menerapkan normalisasi dengan tepat dan mengurangi anomali data.

2.1.4 Kaitan ERD dengan Tujuan Penelitian

Dengan menerapkan ERD dan normalisasi hingga 3NF, dataset akan terstruktur menjadi tabel-tabel relasional yang menghilangkan redundansi (mis. pengulangan atribut gender dan age_group pada setiap transaksi), sehingga memudahkan pemeliharaan data dan analisis selanjutnya.

2.1.5 Analisis Data Transaksi

Analisis data transaksi meliputi pengolahan data pembelian, data pelanggan, metode pembayaran, rating, serta ulasan yang diberikan pengguna dengan tujuan utama memahami pola perilaku pelanggan secara lebih mendalam. Proses analisis ini mencakup pengumpulan dan pengelompokan informasi terkait frekuensi pembelian, jenis produk yang paling diminati, preferensi metode pembayaran, hingga penilaian konsumen terhadap kualitas layanan dan produk. Melalui analisis tersebut, perusahaan dapat mengidentifikasi tren pembelian, mengetahui faktor yang memengaruhi keputusan pelanggan, mengevaluasi kepuasan konsumen, serta menemukan peluang untuk meningkatkan strategi pemasaran. Dengan demikian, hasil analisis data transaksi dapat digunakan sebagai dasar pengambilan keputusan yang lebih efektif, baik untuk meningkatkan layanan, mengoptimalkan penawaran produk, maupun menyusun program loyalitas yang lebih tepat sasaran (Nguyen, T. T., & Simkin, L, 2021).

2.2 Metode Penggerjaan

Metode penggerjaan dalam penelitian ini disusun untuk memastikan bahwa proses analisis, normalisasi, dan perancangan basis data dilakukan secara sistematis, terstruktur, dan sesuai dengan standar perancangan database yang berlaku. Setiap tahapan dirancang agar saling terintegrasi, dimulai dari pengumpulan kebutuhan

sistem, analisis data yang relevan, hingga proses normalisasi untuk menghindari terjadinya redundansi data dan memastikan konsistensi informasi. Setelah itu, dilakukan perancangan model konseptual, logis, dan fisik yang menggambarkan hubungan antar-entitas secara jelas dan terukur. Metode ini juga mencakup proses validasi desain untuk memastikan bahwa struktur basis data yang dibangun mampu mendukung kebutuhan operasional sistem secara optimal. Dengan pendekatan yang sistematis ini, penelitian dapat menghasilkan rancangan basis data yang efisien, mudah dikembangkan, serta mampu mengakomodasi kebutuhan pengguna di masa depan (Han, J., Kamber, M., & Pei, J, 2011).

2.2.1 Pengumpulan dan Observasi Dataset

Langkah awal yang dilakukan adalah mengumpulkan data dari file *ecommerce_dataset_10000.csv*. Dataset tersebut kemudian diamati untuk memahami struktur awalnya, seperti jenis atribut, jumlah kolom, tipe data, serta pengulangan informasi. Observasi ini bertujuan untuk mengidentifikasi adanya redundansi dan potensi anomali yang dapat diatasi melalui proses normalisasi.

2.2.2 Identifikasi Atribut dan Dependensi Fungsional

Setiap kolom dalam dataset dianalisis untuk menentukan dependensi fungsional antara atribut (Connolly, T., & Begg, C, 2015). Tahap ini penting untuk menentukan apakah suatu atribut bergantung pada primary key atau hanya sebagian dari composite key. Hasil identifikasi ini menjadi dasar dalam menentukan bentuk normal (1NF, 2NF, atau 3NF) yang belum terpenuhi.

2.2.3 Penerapan Normalisasi (1NF–3NF)

Proses normalisasi dilakukan secara bertahap sebagai berikut:

1. 1NF: memastikan bahwa semua atribut bersifat atomik dan tidak terdapat pengulangan data dalam satu kolom.
2. 2NF: memisahkan atribut yang mengalami ketergantungan parsial pada primary key.
3. 3NF: menghilangkan ketergantungan transitif agar setiap atribut hanya bergantung

pada primary key.

Hasil dari tahapan ini adalah pemecahan tabel menjadi beberapa entitas baru seperti Customers, Products, Orders, dan Order_Items.

2.2.4 Validasi Struktur Relasional

Desain ERD kemudian divalidasi dengan memeriksa konsistensi relasi antar entitas, kejelasan foreign key, serta penerapan integritas referensial untuk memastikan bahwa keterhubungan data telah sesuai dengan struktur yang direncanakan. Proses validasi ini dilakukan untuk memastikan bahwa hasil normalisasi benar-benar memenuhi standar Third Normal Form (3NF), sehingga setiap atribut berada pada posisi yang tepat dan tidak bergantung pada atribut non-kunci secara berlebihan. Melalui validasi tersebut, desain basis data dipastikan bebas dari berbagai anomali seperti anomali penyisipan, penghapusan, dan pembaruan yang dapat mengganggu konsistensi data di dalam sistem. Dengan demikian, ERD yang telah divalidasi tidak hanya kuat secara struktur tetapi juga andal sebagai fondasi dalam implementasi basis data yang efisien dan stabil.

2.2.5 Implementasi ke RDBMS (Opsional/Penguji Struktur)

Sebagai langkah uji coba, struktur tabel hasil normalisasi dapat diimplementasikan ke dalam Sistem Manajemen Basis Data Relasional (RDBMS) seperti MySQL atau PostgreSQL untuk memastikan bahwa desain yang telah dibuat dapat berfungsi dengan baik pada lingkungan nyata. Implementasi ini memungkinkan peneliti menguji eksekusi berbagai query dasar, termasuk operasi SELECT dengan JOIN, guna memverifikasi apakah relasi antar tabel berjalan sesuai dengan rancangan konseptual dan logis yang telah disusun. Melalui proses ini, dapat diketahui apakah foreign key telah berfungsi dengan benar, apakah data dapat diakses tanpa error, serta apakah integritas referensial tetap terjaga saat dilakukan operasi penyisipan, pembaruan, atau penghapusan. Dengan demikian, tahap implementasi awal ini berperan penting dalam memastikan bahwa basis data tidak hanya benar secara teori tetapi juga stabil dan dapat dioperasikan secara optimal dalam praktik.

BAB 3

METODOLOGI PENELITIAN

3.1 Diagram Penelitian

Diagram penelitian ini menggambarkan alur sistematis yang digunakan dalam membangun struktur basis data dari dataset, mulai dari tahap pengumpulan data hingga implementasi akhir. Proses penelitian dimulai dengan pengumpulan dataset, yaitu memperoleh file mentah yang menjadi dasar analisis. Setelah dataset diperoleh, langkah berikutnya adalah observasi struktur data, di mana seluruh kolom dan isi dataset ditinjau untuk mengidentifikasi atribut utama, pola redundansi, ketergantungan antar data, serta potensi anomali.

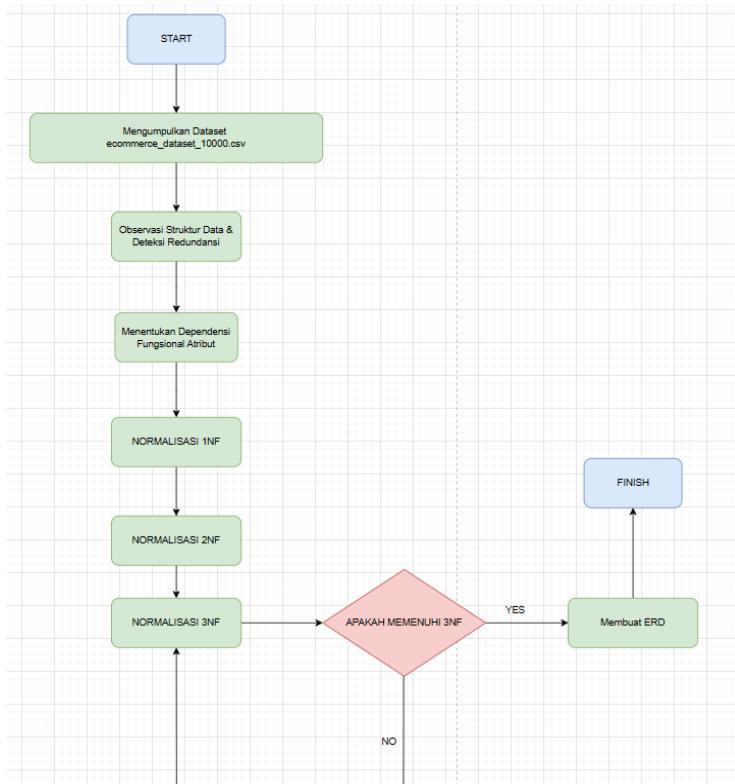
Tahap berikutnya adalah analisis dependensi fungsional, yang bertujuan menentukan hubungan logis antar atribut sehingga proses normalisasi dapat dilakukan dengan benar. Setelah dependensi ditetapkan, dilakukan proses normalisasi secara bertahap mulai dari 1NF, 2NF, hingga 3NF. Tahap normalisasi ini memastikan bahwa struktur data bebas dari duplikasi, ketergantungan parsial, dan ketergantungan transitif. Selama proses ini, dataset yang awalnya berada dalam satu tabel besar dipecah menjadi empat entitas terpisah: Customers, Products, Orders, dan Order_Items.

Hasil normalisasi kemudian digunakan untuk menyusun ERD (Entity Relationship Diagram) yang menggambarkan hubungan antar tabel secara visual dan struktural. Setelah ERD selesai, dilakukan validasi relasi untuk memastikan bahwa setiap foreign key terhubung secara konsisten dan tidak terjadi pelanggaran integritas referensial.

Tahap berikutnya adalah implementasi struktur ke MySQL, yang melibatkan pembuatan tabel dengan perintah *CREATE TABLE*, pembersihan data dari error seperti #REF! maupun format yang tidak valid, serta penyesuaian tipe data untuk mengatasi error seperti 1366, 1110, dan tanggal 0000-00-00. Setelah data berhasil di-*import* tanpa error, dilakukan uji *query JOIN* untuk memastikan relasi antar tabel berjalan sebagaimana mestinya.

Alur penelitian ditutup dengan tahap penyusunan hasil dan pembahasan, di mana seluruh temuan, struktur tabel final, ERD, serta kendala teknis selama proses implementasi dijelaskan secara terperinci sebagai bagian dari laporan akhir.

3.1.1 Flowchart



Gambar 3.1 Flowchart Penelitian

Flowchart penelitian ini menggambarkan alur kerja dalam merancang basis data dari dataset *ecommerce_dataset_10000.csv*, dimulai dari tahap START sebagai penanda awal proses. Kegiatan pertama yang dilakukan adalah mengumpulkan dataset, yang menjadi dasar analisis keseluruhan penelitian. Selanjutnya, dilakukan observasi struktur data dan deteksi redundansi untuk mengidentifikasi pola pengulangan data, ketidakkonsistenan atribut, serta potensi anomali. Tahap berikutnya adalah menentukan dependensi fungsional atribut, yang menjadi dasar untuk melakukan normalisasi secara benar. Proses normalisasi dilaksanakan bertahap mulai dari 1NF, 2NF, hingga 3NF, masing-masing bertujuan menghilangkan data tidak atomik, menghapus ketergantungan parsial, dan menghilangkan ketergantungan transitif. Setelah mencapai 3NF, dilakukan evaluasi melalui proses decision “Apakah Memenuhi 3NF?”; apabila struktur belum sesuai, proses normalisasi diulang, namun jika telah memenuhi 3NF maka dilanjutkan ke tahap pembuatan ERD sebagai representasi visual relasi antar entitas hasil normalisasi. Proses penelitian ditutup pada tahap FINISH, menandai bahwa struktur basis data telah berhasil dibangun secara sistematis dan siap diimplementasikan lebih lanjut.

3.2 Jadwal Penelitian

Jadwal penelitian disusun untuk menggambarkan alur waktu pelaksanaan setiap tahap, mulai dari pengumpulan dataset, observasi struktur data, analisis dependensi fungsional, dan normalisasi hingga 3NF. Setelah struktur data dinyatakan valid, kegiatan dilanjutkan dengan penyusunan ERD, implementasi skema database ke MySQL, serta penyusunan hasil dan laporan akhir. Seluruh tahapan direncanakan secara sistematis agar penelitian berjalan terarah dan selesai tepat waktu.

Tabel 3.1 Jadwal Pelaksanaan Penelitian

BAB 4

HASIL

4.1 Pengertian

Bagian hasil dalam penelitian ini menjelaskan output yang diperoleh setelah dilakukan proses observasi dataset, identifikasi atribut, normalisasi, serta penyusunan struktur basis data menggunakan ERD sesuai rancangan akhir. Dataset yang awalnya berada dalam satu tabel besar menghasilkan beberapa temuan penting setelah melalui proses normalisasi hingga Bentuk Normal Ketiga (3NF).

4.1.1 Struktur Data Mentah yang Telah Dianalisis

Dataset awal yang terdiri dari data pelanggan, data produk, data transaksi, serta rating dan review pelanggan dianalisis untuk menentukan pola hubungan antar atribut. Dari proses ini ditemukan adanya pengulangan data seperti gender, age_group, dan country yang muncul berulang setiap kali pelanggan melakukan transaksi. Temuan ini menguatkan perlunya pemecahan tabel agar struktur data lebih efisien.

4.1.2 Hasil Pemecahan Tabel melalui Normalisasi (1NF – 3NF)

Proses normalisasi pada dataset *sbd kelompok 3.xlsx* menghasilkan pemecahan tabel yang lebih terstruktur dan efisien. Pada tahap awal (1NF), dataset yang semula berada dalam satu tabel besar diperiksa untuk memastikan bahwa setiap kolom berisi nilai atomik dan tidak terdapat data ganda dalam satu sel. Pada tahap ini dilakukan pembersihan terhadap kolom-kolom yang memuat nilai berulang serta penyelarasan format data agar setiap atribut memiliki satu nilai yang jelas dan konsisten. Setelah struktur dasar terpenuhi, proses dilanjutkan ke tahap 2NF, yaitu menghilangkan ketergantungan parsial antara atribut non-key dengan sebagian primary key. Karena dataset awal tidak menggunakan composite key, fokus pada tahap ini adalah memisahkan atribut-atribut yang tidak terkait langsung dengan transaksi, seperti informasi pelanggan dan detail produk. Pemisahan ini menghasilkan tabel Customers yang menyimpan informasi personal pelanggan (gender, age_group, country,

signup_date) dan tabel Products yang berisi informasi produk (category, unit_price, product_name).

Tahap berikutnya adalah normalisasi 3NF, yang bertujuan menghilangkan ketergantungan transitif antara atribut non-key. Pada tahap ini, dipastikan bahwa setiap atribut non-primary key bergantung langsung pada primary key masing-masing entitas, bukan pada atribut lain. Hasilnya, atribut yang terkait dengan transaksi pembelian dipindahkan ke tabel Orders, yang mencakup order_id, customer_id, product_id, order_date, payment_method, dan order_status. Sementara itu, informasi tambahan yang bersifat detail item transaksi seperti rating, review_text, dan review_date dipisahkan ke tabel Reviews/Order_Items agar tidak bercampur dengan data transaksi utama.

Melalui ketiga tahap normalisasi tersebut, dataset berhasil dipisahkan menjadi empat tabel inti yang saling berhubungan melalui foreign key namun tidak lagi mengalami redundansi: Customers, Products, Orders, dan Reviews/Order_Items. Struktur baru ini memungkinkan analisis dilakukan lebih efisien, menjaga integritas data, dan meminimalkan anomali saat proses insert, update, maupun delete dilakukan di dalam sistem manajemen basis data.

Tabel 4.1 Tabel 1NF

First name	Last name	Gender	age_group	signup_date	country	product_id	product_name
Erica	Oliver	Female	Teenagers	6/29/2022	Canada	PROD108	Fitbit Versa 3
Christopher	White	Male	Adults	8/24/2023	China	PROD103	Levi's Jeans
Spencer	Foster	Male	Senior	7/18/2023	Germany	PROD111	Lego Star Wars Set
Jessica	Harris	Male	Teenagers	8/22/2025	France	PROD107	Dyson Vacuum
Amy	Johnson	Female	Teenagers	3/23/2021	Brazil	PROD105	Adidas Running Shoes
Shelby	Sutton	Other	Adults	7/18/2025	Canada	PROD108	Fitbit Versa 3
Jennifer	Baxter	Female	Senior	11/30/2023	France	PROD106	Instant Pot

Tabel 4.2 Tabel 2NF

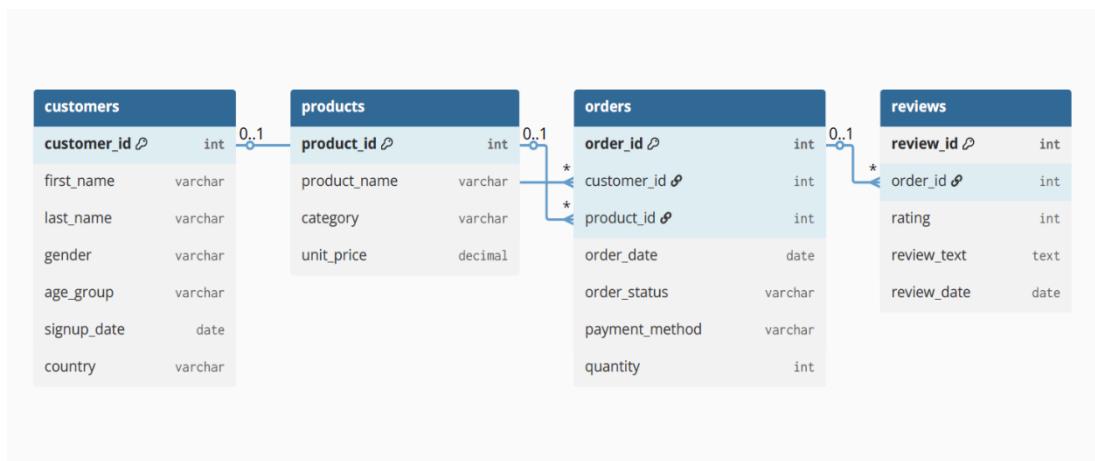
Age group	signup_date	country	product_id	product_name	category	quantity	unit_price
Teenagers	6/29/2022	Canada	PROD108	Fitbit Versa 3	Electronics	3	229
Adults	8/24/2023	China	PROD103	Levi's Jeans	Apparel	4	59
Senior	7/18/2023	Germany	PROD111	Lego Star Wars Set	Toys	2	59
Teenagers	8/22/2025	France	PROD107	Dyson Vacuum	Home & Kitchen	4	399
Teenagers	3/23/2021	Brazil	PROD105	Adidas Running Shoes	Apparel	1	110
Adults	7/18/2025	Canada	PROD108	Fitbit Versa 3	Electronics	5	229
Senior	11/30/2023	France	PROD106	Instant Pot	Home & Kitchen	1	99

Tabel 4.3 Tabel 3NF

product_id	product_name	category	quantity	unit_price
PROD108	Fitbit Versa 3	Electronics	3	229
PROD103	Levi's Jeans	Apparel	4	59
PROD111	Lego Star Wars Set	Toys	2	59
PROD107	Dyson Vacuum	Home & Kitchen	4	399
PROD105	Adidas Running Shoes	Apparel	1	110
PROD108	Fitbit Versa 3	Electronics	5	229
PROD106	Instant Pot	Home & Kitchen	1	99

4.1.3 Hasil Penyusunan ERD (*Entity-Relationship Diagram*)

Entity Relationship Diagram (ERD) yang dihasilkan menunjukkan struktur basis data yang telah dinormalisasi menjadi empat entitas utama, yaitu customers, products, orders, dan reviews, yang saling terhubung melalui relasi *one-to-many*. Tabel customers menyimpan informasi profil pelanggan dengan *customer_id* sebagai primary key yang terhubung ke tabel orders, di mana satu pelanggan dapat memiliki banyak transaksi. Tabel products menyimpan data produk dengan *product_id* sebagai primary key, yang juga dihubungkan ke tabel orders karena satu produk dapat muncul pada banyak transaksi. Tabel orders menjadi pusat transaksi dengan *order_id* sebagai primary key serta foreign key *customer_id* dan *product_id* yang menghubungkannya dengan pelanggan dan produk. Sementara itu, tabel reviews menyimpan ulasan pelanggan dengan *review_id* sebagai primary key dan *order_id* sebagai foreign key, sehingga satu order dapat memiliki nol atau satu review. Keseluruhan ERD ini menggambarkan hubungan logis dan terstruktur yang mencerminkan proses bisnis e-commerce: pelanggan melakukan pembelian produk melalui sebuah order, kemudian dapat memberikan review, dengan setiap tabel memiliki peran yang jelas dan bebas dari redundansi.



Gambar 4.1 Hasil Penyusunan ERD (*Entity-Relationship Diagram*)

4.1.4 Hasil Akhir Struktur Basis Data Relasional

Dengan menerapkan tahapan penggeraan secara sistematis, hasil akhir penelitian berupa struktur database relasional yang efisien, bebas redundansi, dan siap digunakan untuk analisis lanjutan seperti analisis pembelian, rating pelanggan, atau implementasi dashboard.

4.2 Hasil Uji Query

Pada tahap ini dilakukan pengujian query untuk memastikan bahwa struktur basis data hasil normalisasi telah berfungsi dengan baik dan seluruh relasi antar tabel dapat berjalan sesuai rancangan ERD. Uji query dilakukan setelah proses *import* data ke MySQL dan setelah seluruh error terkait tipe data, format tanggal, serta duplikasi kolom berhasil diatasi:

1. Siapa pasangan produk yang sering dibeli customer yang sama (tanpa limit)?

Query ini mencari pasangan produk yang sering dibeli oleh customer yang sama. Teknik yang digunakan adalah self join pada tabel *orders* sehingga setiap transaksi customer bisa dicocokkan dengan transaksi lainnya yang dilakukan oleh customer yang sama. Syarat *product_id < product_id* digunakan untuk mencegah duplikasi pasangan seperti (A, B) dan (B, A). Setelah itu, jumlah kemunculan setiap pasangan dihitung, lalu diurutkan dari yang paling sering.

Query;

```
SELECT
    LEAST(o1.product_id, o2.product_id) AS product_1,
    GREATEST(o1.product_id, o2.product_id) AS product_2,
    COUNT(*) AS pair_count
FROM orders o1
JOIN orders o2
    ON o1.customer_id = o2.customer_id
    AND o1.product_id < o2.product_id
GROUP BY product_1, product_2
ORDER BY pair_count DESC;
```

produk_a	produk_b	jumlah_kemunculan
PROD107	PROD110	2
PROD100	PROD109	2
PROD103	PROD112	2
PROD106	PROD108	2
PROD106	PROD109	1
PROD102	PROD109	1
PROD101	PROD106	1
PROD103	PROD111	1
PROD100	PROD104	1
PROD100	PROD103	1
PROD109	PROD110	1
PROD101	PROD114	1
PROD110	PROD113	1
PROD100	PROD101	1
PROD102	PROD114	1
PROD111	PROD113	1
PROD109	PROD113	1

Gambar 4.2 Hasil Query pasangan produk yang sering dibeli customer yang sama

2. Kategori produk apa yang memberi total omzet tertinggi?

Query ini menghitung total omzet setiap kategori produk dengan cara menjumlahkan hasil perkalian antara quantity dari tabel *orders* dan unit_price dari tabel *products*. Data order disambungkan dengan data produk berdasarkan product_id untuk mendapatkan harga dan kategori produk. Nilai omzet ini kemudian dijumlahkan per kategori dan diurutkan dari yang terbesar.

Query:

```
SELECT
    p.category,
    SUM(o.quantity * p.unit_price) AS total_omzet
FROM orders o
JOIN products p ON o.product_id = p.product_id
GROUP BY p.category
ORDER BY total_omzet DESC;
```

category	total_omzet
Electronics	275323.00
Home & Kitchen	57384.00
Apparel	31721.00
Toys	25689.00
Books	18882.00
Sports	16183.00

Gambar 4.3 Hasil Query produk yang memberi total omzet tertinggi

3. Negara mana yang memberikan total omzet paling besar?

Query ini mencari negara yang memberikan omzet terbesar dengan menggabungkan tabel *orders* dengan *customers* untuk mengetahui negara customer, serta tabel *products* untuk mendapatkan harga produk. Omzet dihitung menggunakan rumus quantity × unit_price, lalu dijumlahkan per negara. Hasilnya menunjukkan negara yang memberikan pemasukan tertinggi.

Query:

```
SELECT
    c.country,
```

```

        SUM(o.quantity * p.unit_price) AS total_omzet
    FROM orders o
    JOIN customers c ON o.customer_id = c.customer_id
    JOIN products p ON o.product_id = p.product_id
    GROUP BY c.country
    ORDER BY total_omzet DESC;

```

country	total_omzet
China	56674.00
India	51415.00
USA	50344.00
France	44278.00
Japan	43401.00
Brazil	41450.00
Canada	35427.00
Germany	34950.00
UK	34475.00
Australia	32768.00

Gambar 4.4 Hasil Query Negara yang memberikan total omzet paling besar

4. Metode pembayaran mana yang mendapatkan rata-rata rating tertinggi? Berapa rata-rata rating serta jumlah review untuk setiap metode pembayaran yang digunakan di tabel orders?

Query ini menentukan metode pembayaran yang memiliki rating rata-rata tertinggi serta menghitung jumlah review untuk masing-masing metode. Dengan menggunakan LEFT JOIN, order yang tidak memiliki review tetap dihitung. Rata-rata rating dihitung menggunakan AVG, sedangkan jumlah review dihitung menggunakan COUNT. Hasil diurutkan berdasarkan rating tertinggi.

Query:

```

SELECT
    o.payment_method,
    AVG(r.rating) AS avg_rating,
    COUNT(r.review_id) AS total_reviews
FROM orders o
LEFT JOIN reviews r ON o.order_id = r.order_id
GROUP BY o.payment_method
ORDER BY avg_rating DESC;

```

payment_method	avg_rating	total_review
Cash on Delivery	3.1775	169
Credit Card	3.0203	197
PayPal	2.9106	179

Gambar 4.5 Hasil Query rata rata rating serta jumlah review

5. Produk mana yang paling banyak menghasilkan omzet berdasarkan jumlah pembelian \times unit_price?

Query ini menghitung omzet per produk dengan mengalikan quantity dengan unit_price. Semua nilai omzet dihitung untuk tiap produk dan dijumlahkan. Nilai total omzet kemudian diurutkan dari yang terbesar untuk mengetahui produk mana yang menghasilkan pemasukan paling besar.

Query:

```
SELECT
    p.product_id,
    p.product_name,
    SUM(o.quantity * p.unit_price) AS total_omzet
FROM orders o
JOIN products p ON o.product_id = p.product_id
GROUP BY p.product_id, p.product_name
ORDER BY total_omzet DESC;
```

product_id	product_name	total_omzet
PROD101	Samsung Galaxy S23	117769.00
PROD100	iPhone 14	103896.00
PROD107	Dyson Vacuum	47880.00
PROD102	Sony Headphones	26865.00
PROD108	Fitbit Versa 3	26793.00
PROD112	Barbie Dreamhouse	20497.00
PROD104	Nike Air Max	13920.00
PROD113	Wilson Tennis Racket	13708.00
PROD109	Kindle Paperwhite	12642.00
PROD105	Adidas Running Shoes	10780.00
PROD106	Instant Pot	9504.00
PROD103	Levi's Jeans	7021.00
PROD110	Harry Potter Box Set	6240.00
PROD111	Lego Star Wars Set	5192.00
PROD114	Yoga Mat	2475.00

Gambar 4.6 Hasil Query produk yang paling banyak menghasilkan omzet

6. Negara yang gender nya paling beragam?

Query ini menghitung jumlah gender unik di setiap negara dengan menggunakan COUNT DISTINCT pada kolom gender. Setiap negara dihitung variasi gendernya, lalu diurutkan dari negara yang paling beragam.

Query:

```
SELECT
    country,
    COUNT(DISTINCT gender) AS gender_variety
FROM customers
GROUP BY country
ORDER BY gender_variety DESC;
```

country	variasi_gender
Australia	3
Brazil	3
Canada	3
China	3
France	3
Germany	3
India	3
Japan	3
UK	3
USA	3
country	1

Gambar 4.7 Hasil Query Negara yang gendernya paling beragam

7. Negara manakah yang memiliki jumlah pelanggan aktif (pernah order) terbanyak?

Query ini menentukan negara yang memiliki jumlah pelanggan aktif terbanyak, yaitu pelanggan yang pernah melakukan order. Dengan melakukan JOIN antara *orders* dan *customers*, kemudian menghitung jumlah customer unik yang pernah melakukan transaksi, hasilnya menunjukkan negara dengan pelanggan aktif

terbanyak.

Query:

```
SELECT
    c.country,
    COUNT(DISTINCT o.customer_id) AS active_customers
FROM orders o
JOIN customers c ON o.customer_id = c.customer_id
GROUP BY c.country
ORDER BY active_customers DESC;
```

country	total_active_customer
China	60
Brazil	60
France	56
UK	53
India	52
USA	51
Canada	50
Germany	48
Japan	47
Australia	45

Gambar 4.8 Hasil Query Negara yang memiliki jumlah pelanggan aktif terbanyak

8. Produk manakah yang paling banyak dibeli tapi paling sedikit direview?

Query ini mencari produk yang memiliki volume pembelian tinggi tetapi jumlah review rendah. Total pembelian dihitung dari sum(quantity) dan review dihitung dari jumlah review_id. Produk kemudian diurutkan berdasarkan jumlah pembelian tertinggi dan review paling sedikit.

Query:

```
SELECT
    p.product_id,
    p.product_name,
    SUM(o.quantity) AS total_quantity_sold,
    COUNT(r.review_id) AS total_reviews
FROM products p
JOIN orders o ON p.product_id = o.product_id
LEFT JOIN reviews r ON o.order_id = r.order_id
GROUP BY p.product_id, p.product_name
ORDER BY total_quantity_sold DESC, total_reviews ASC;
```

product_id	product_name	unique_buyers	total_reviews	selisih ▾ 1
PROD100	iPhone 14	39	39	0
PROD101	Samsung Galaxy S23	43	43	0
PROD102	Sony Headphones	46	46	0
PROD104	Nike Air Max	35	35	0
PROD105	Adidas Running Shoes	29	29	0
PROD106	Instant Pot	34	34	0
PROD107	Dyson Vacuum	39	39	0
PROD108	Fitbit Versa 3	39	39	0
PROD109	Kindle Paperwhite	36	36	0
PROD110	Harry Potter Box Set	32	32	0
PROD111	Lego Star Wars Set	31	31	0
PROD112	Barbie Dreamhouse	37	37	0
PROD114	Yoga Mat	33	33	0
product_id (PK)	product_name	0	0	0
PROD103	Levi's Jeans	39	40	-1
PROD113	Wilson Tennis Racket	31	32	-1

Gambar 4.9 Produk yang paling banyak dibeli tapi paling sedikit direview

9. Apa Tiga barang spesifik apa yang jadi *Best Seller*?

Query ini mencari tiga produk yang memiliki total pembelian terbanyak. Pembelian dihitung dengan menjumlahkan quantity per produk, kemudian diurutkan dari yang paling besar, dan hanya tiga teratas ditampilkan menggunakan LIMIT 3.

Query:

```
SELECT
    p.product_id,
    p.product_name,
    SUM(o.quantity) AS total_sold
FROM products p
JOIN orders o ON p.product_id = o.product_id
GROUP BY p.product_id, p.product_name
ORDER BY total_sold DESC
LIMIT 3;
```

product_id	product_name	total_units_sold ▾ 1
PROD102	Sony Headphones	135
PROD101	Samsung Galaxy S23	131
PROD107	Dyson Vacuum	120

Gambar 4.10 Hasil Query tiga produk yang jadi *Best Seller*?

10. Kategori manakah yang paling besar variasi harganya (harga tertinggi – harga terendah)?

Query terakhir mencari kategori produk yang memiliki rentang harga terbesar dengan menghitung selisih antara harga tertinggi (MAX) dan harga terendah (MIN) dalam kategori yang sama. Hasilnya diurutkan dari variasi harga terbesar.

Query:

```
SELECT
    category,
    (MAX(unit_price) - MIN(unit_price)) AS price_variation
FROM products
GROUP BY category
ORDER BY price_variation DESC;
```

	<input type="button" value="Ubah"/>	<input type="button" value="Salin"/>	<input type="button" value="Hapus"/>	category	price_range
<input type="checkbox"/>	<input type="button" value="Ubah"/>	<input type="button" value="Salin"/>	<input type="button" value="Hapus"/>	Electronics	800.00
<input type="checkbox"/>	<input type="button" value="Ubah"/>	<input type="button" value="Salin"/>	<input type="button" value="Hapus"/>	Home & Kitchen	300.00
<input type="checkbox"/>	<input type="button" value="Ubah"/>	<input type="button" value="Salin"/>	<input type="button" value="Hapus"/>	Toys	140.00
<input type="checkbox"/>	<input type="button" value="Ubah"/>	<input type="button" value="Salin"/>	<input type="button" value="Hapus"/>	Sports	124.00
<input type="checkbox"/>	<input type="button" value="Ubah"/>	<input type="button" value="Salin"/>	<input type="button" value="Hapus"/>	Apparel	61.00
<input type="checkbox"/>	<input type="button" value="Ubah"/>	<input type="button" value="Salin"/>	<input type="button" value="Hapus"/>	Books	49.00
<input type="checkbox"/>	<input type="button" value="Ubah"/>	<input type="button" value="Salin"/>	<input type="button" value="Hapus"/>	category	0.00

Gambar 4.11 Hasil Query produk yang paling besar variasi harganya

BAB 5

PENUTUP

5.1 Kesimpulan

Berdasarkan rangkaian proses penelitian yang telah dilakukan, dapat disimpulkan bahwa dataset yang pada awalnya disusun dalam satu tabel besar mengandung sejumlah permasalahan, seperti redundansi data, ketergantungan atribut yang tidak tepat, serta potensi anomali pada operasi *insert*, *update*, dan *delete*. Melalui observasi struktur data, analisis dependensi fungsional, dan penerapan normalisasi hingga Bentuk Normal Ketiga (3NF), dataset berhasil diubah menjadi struktur basis data yang lebih rapi, efisien, dan bebas duplikasi.

Proses normalisasi menghasilkan empat entitas utama, yaitu Customers, Products, Orders, dan Order_Items, yang kemudian divisualisasikan dalam bentuk Entity Relationship Diagram (ERD). ERD ini menjadi pedoman dalam membangun kembali skema basis data secara relasional dan memastikan bahwa setiap entitas memiliki hubungan yang jelas dan memenuhi prinsip integritas referensial.

Implementasi ke dalam MySQL juga memberikan berbagai temuan teknis yang penting. Beberapa error seperti Error 1366, format tanggal 0000-00-00, dan Error 1110 berhasil diatasi melalui penyesuaian tipe data, pembersihan struktur file CSV, serta pengubahan format input. Dengan penyelesaian kendala tersebut, database berhasil dibangun dan dapat diuji menggunakan berbagai query, termasuk *JOIN*, tanpa mengalami konflik relasi.

Secara keseluruhan, penelitian ini membuktikan bahwa penerapan normalisasi dan perancangan ERD sangat penting dalam memastikan struktur basis data dapat digunakan secara optimal, mudah dianalisis, serta siap dikembangkan lebih lanjut untuk kebutuhan pemrosesan data maupun pengembangan aplikasi.

5.2 Saran

Untuk meningkatkan kualitas dan pengembangan basis data di masa mendatang, beberapa saran dapat diterapkan. Pertama, proses validasi data perlu dilakukan secara berkala agar error seperti *incorrect value*, format tanggal tidak valid, maupun data kosong dapat dicegah sejak awal sebelum data diimpor ke sistem. Kedua,

pemilihan tipe data yang fleksibel harus dipertimbangkan sejak tahap perancangan, misalnya penggunaan tipe VARCHAR untuk kolom ID atau tanggal guna menangani data yang tidak konsisten. Ketiga, dokumentasi proses normalisasi dan ERD perlu disimpan dengan baik karena akan sangat membantu ketika database perlu diperluas, dimodifikasi, atau digunakan oleh tim lain. Selain itu, penambahan indeks pada kolom penting seperti *customer_id*, *product_id*, dan *order_id* dianjurkan agar kinerja query semakin optimal, terutama bila ukuran dataset bertambah besar. Database yang telah terbentuk juga berpotensi untuk dikembangkan lebih lanjut ke dalam aplikasi, seperti dashboard analisis penjualan atau sistem manajemen transaksi. Selanjutnya, apabila data sudah konsisten, penggunaan tipe data tanggal sebaiknya dievaluasi ulang dan dikembalikan ke tipe DATE atau DATETIME agar fungsi waktu dapat berjalan lebih optimal. Terakhir, uji integritas relasi perlu dilakukan setelah proses impor data untuk memastikan tidak ada *foreign key* yang rusak akibat kesalahan pada file CSV.

DAFTAR PUSTAKA

- Laudon, K. C., & Traver, C. G. (2022). *E-commerce trends and technologies*. Retrieved from
https://www.researchgate.net/publication/361709659_E-Commerce_Trends_and_Technologies
- Nguyen, T. T., & Simkin, L. (2021). Understanding customer behavior in digital commerce. *Journal of Retailing and Consumer Services*, 63, 102735.
<https://doi.org/10.1016/j.jretconser.2021.102735>
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques*. Retrieved from
<https://books.google.com/books?id=pQws07tdpJoC>
- Connolly, T., & Begg, C. (2015). *Database Systems: A Practical Approach to Design, Implementation, and Management*. Pearson.
- W3Schools. (2024). *SQL Database basics*. Retrieved from
https://www.w3schools.com/sql/sql_intro.asp