JUMABI : Jurnal Manajemen, Akuntansi dan Bisnis

e-ISSN: 2985-3834 (Online - Elektronik)
Volume.3 | Nomor.2 | Juli | 2025 | PP : 67-77
DOI: https://doi.org/10.56314/jumabi.v3i2.332

Mengatasi Duplikasi Data Dalam Database Menggunakan Metode Eliminasi Data

Dinda Puspita Sari[1*], Septia Amelia[2], Nadhifa Salsabila Bakris[3] Nurbaiti[4]
[1,2,3,4]Program Studi Manajemen, Fakultas Ekonomi Dan Bisnis Islam Universitas Islam Negeri Sumatera Utara, Indonesia

Corresponding author*
E-mail:
dindapuspita2303@gmail.com

Article history
Received *: 2025-06-23*
Revised *: 2025-07-05*
Accepted *: 2025-07-12*

Copyright © 2025 Author.
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Abstrak

Dalam era digital yang terus berkembang, duplikasi data dalam database menjadi tantangan serius yang berdampak pada efisiensi penyimpanan, keakuratan informasi, dan kualitas pengambilan keputusan. Penelitian ini bertujuan untuk mengatasi permasalahan duplikasi data melalui penerapan metode eliminasi data. Metode ini dilakukan melalui tahapan sistematis, dimulai dari pengumpulan data, identifikasi duplikasi berdasarkan atribut tertentu, hingga penghapusan data rangkap secara permanen menggunakan pendekatan pencocokan sederhana dan pemanfaatan bahasa pemrograman SQL. Proses eliminasi data dilakukan untuk memastikan bahwa setiap entri dalam database bersifat unik dan valid, dengan tujuan meningkatkan integritas dan keandalan sistem informasi. Hasil dari penelitian ini menunjukkan bahwa metode eliminasi data dapat secara efektif mengurangi redundansi, meningkatkan efisiensi pengelolaan database, dan mendukung keputusan berbasis data yang lebih akurat. Dengan demikian, implementasi teknik eliminasi data menjadi langkah strategis dalam menjaga kualitas sistem informasi di berbagai sektor.

Kata kunci: Eliminasi Data, Duplikasi, Database, SQL, Pencocokan Sederhana, Kualitas Informasi.

PENDAHULUAN

Data merupakan aset berharga dalam era digital yang terus berkembang. Dalam berbagai sektor, seperti bisnis, pemerintahan, kesehatan, dan pendidikan, data memainkan peran penting dalam pengambilan keputusan, analisis tren,

serta peningkatan efisiensi operasional (Fadly & Yusuf, 2024). Pengelolaan data yang baik memungkinkan organisasi untuk memperoleh wawasan yang lebih akurat, meningkatkan layanan, serta mengoptimalkan penggunaan sumber daya. Dengan data yang terstruktur dan berkualitas, proses bisnis dapat berjalan lebih lancar dan menghasilkan keputusan yang lebih tepat.Namun, dalam praktiknya, salah satu tantangan utama dalam pengelolaan data adalah terjadinya duplikasi data dalam database. Duplikasi data dapat menyebabkan berbagai permasalahan, seperti inkonsistensi informasi, redundansi penyimpanan, serta kesalahan dalam analisis dan pengambilan keputusan. Penyebab utama duplikasi data meliputi kesalahan input manual, penggabungan data dari berbagai sumber tanpa validasi yang memadai, serta kurangnya mekanisme pencocokan dan verifikasi data yang efektif. (Wali,Efitra et al., 2023)

Dampak dari duplikasi data sangat signifikan, terutama dalam sistem yang bergantung pada keakuratan informasi. Misalnya, dalam dunia bisnis, duplikasi data pelanggan dapat menyebabkan kesalahan dalam strategi pemasaran dan layanan pelanggan. Di sektor kesehatan, duplikasi data pasien dapat mengakibatkan kesalahan dalam diagnosis dan perawatan medis. Oleh karena itu, diperlukan metode yang efektif untuk mendeteksi dan mengatasi duplikasi data guna memastikan integritas dan keandalan informasi dalam database.Duplikasi data dapat terjadi akibat berbagai faktor, seperti kesalahan input manual, integrasi dari berbagai sumber data yang berbeda, serta kurangnya mekanisme validasi yang efektif. Misalnya, dalam sistem e-commerce, pelanggan yang melakukan registrasi lebih dari satu kali dengan data yang sedikit berbeda dapat menyebabkan kesalahan dalam pengiriman barang.
Begitu pula dalam sistem rumah sakit, duplikasi data pasien dapat mengakibatkan kesalahan dalam diagnosis atau perawatan.Untuk mengatasi permasalahan ini, diperlukan metode yang dapat mengidentifikasi dan mengeliminasi duplikasi secara efektif. Salah satu pendekatan yang dapat digunakan adalah metode pencocokan sederhana (simple matching). Metode ini bertujuan untuk mendeteksi dan menghapus entri data yang memiliki tingkat kemiripan tertentu berdasarkan aturan yang telah ditentukan. Dengan menerapkan eliminasi data , proses identifikasi duplikasi dapat dilakukan secara lebih efisien dan sistematis, sehingga meningkatkan kualitas data dalam database.Hal ini sebagai bukti bahwa databse meemgang peranan penting dalam pengolahan data,diuplikasi bukanlah hal yang bisa di sepelekan baik dalam segala sektor (Rismaninda,Radifa et al ,2025)

Dalam penelitian ini, akan dibahas penerapan metode pencocokan sederhana dalam mengatasi duplikasi data dalam database. Selain itu, akan dievaluasi efektivitas metode ini dalam meningkatkan efisiensi dan akurasi sistem pengelolaan data. Diharapkan penelitian ini dapat memberikan solusi yang bermanfaat dalam menjaga kualitas dan integritas data pada berbagai sistem informasi.

Tinjauan Teoretis

Duplikasi Data

Duplikasi data adalah kondisi di mana terdapat lebih dari satu salinan data yang sama atau sangat mirip dalam sebuah database atau sistem penyimpanan data. Hal ini biasanya terjadi ketika satu entitas atau informasi tercatat berulang kali dalam tabel atau file yang sama, sehingga menghasilkan baris data yang identik atau hampir identik pada beberapa kolom tertentu. Duplikasi data bisa disebabkan oleh berbagai faktor, seperti kesalahan input data, penggabungan data dari sumber yang berbeda tanpa proses penyaringan yang tepat, atau sistem database yang tidak terstruktur dengan baik. Keberadaan data duplikat dapat menimbulkan berbagai masalah, termasuk inkonsistensi informasi, pemborosan ruang penyimpanan, dan ketidakakuratan dalam analisis data atau pelaporan. Misalnya, dalam pengolahan data menggunakan SQL, duplikasi sering muncul ketika operasi JOIN dilakukan dengan kondisi yang tidak lengkap, sehingga menghasilkan baris ganda yang sebenarnya merepresentasikan entitas yang sama. Oleh karena itu, pengelolaan duplikasi data menjadi penting untuk menjaga integritas, keakuratan, dan efisiensi database. Berbagai teknik seperti penggunaan *primary key*, validasi input, dan proses eliminasi data (deduplikasi) diterapkan untuk mengidentifikasi dan menghapus data yang berulang agar database tetap bersih dan dapat diandalkan. Pendekatan deduplikasi ini juga membantu mengurangi kebutuhan penyimpanan dan meningkatkan performa sistem secara keseluruhan. (Tantra 2012)

Database

Database adalah kumpulan data atau informasi yang disimpan secara sistematis dan terorganisir berdasarkan ketentuan tertentu agar saling berhubungan dan mudah dikelola. Data dalam database umumnya disusun dalam bentuk tabel yang terdiri dari baris (*record*) dan kolom *(field*), di mana setiap baris merepresentasikan satu unit data dan setiap kolom menunjukkan jenis atribut tertentu, seperti nama, tanggal, atau nilai numerik. Struktur ini memungkinkan proses pencarian, pengolahan, dan analisis data dilakukan dengan lebih cepat dan efisien. Pengelolaan database dilakukan melalui perangkat lunak khusus yang disebut *Database Management System (DBMS),* yang berfungsi sebagai antarmuka antara pengguna dan database untuk memungkinkan berbagai operasi seperti pencarian, penyimpanan, pembaruan, penghapusan, hingga pengelolaan hak akses pengguna secara efisien dan aman. DBMS juga memiliki kemampuan untuk menjaga integritas data, memastikan konsistensi antar entitas, dan melindungi informasi dari akses tidak sah melalui sistem otentikasi dan kontrol keamanan. (Intan & Irwan, 2024)

Kualitas Data

Kualitas data adalah tingkat keakuratan, konsistensi, kelengkapan, ketepatan waktu, validitas, dan keunikan data yang digunakan dalam suatu

sistem atau proses pengambilan keputusan. Data yang berkualitas tinggi harus mencerminkan kondisi sebenarnya (akurasi), memiliki informasi yang lengkap tanpa kekosongan (kelengkapan), bersifat seragam di seluruh sistem (konsistensi), diperbarui secara tepat waktu (ketepatan waktu), sesuai dengan format dan standar yang ditentukan (validitas), serta bebas dari duplikasi (keunikan). Kualitas data yang buruk dapat menyebabkan kesalahan dalam analisis, pengambilan keputusan yang keliru, pemborosan sumber daya, serta penurunan kepercayaan pengguna terhadap sistem informasi (Susilo & Yudhogiri 2020). Mereka juga mengemukakan bahwa penerapan kerangka Total Data Quality Management (TDQM) dapat membantu institusi dalam memetakan dimensi kualitas data dan melakukan perbaikan secara berkelanjutan.

Eliminasi Data

Eliminasi data, merupakan serangkaian tindakan terstruktur dan terencana untuk permanen menghapus data yang telah mencapai akhir siklus hidupnya atau dianggap tidak lagi memiliki nilai bisnis, operasional, hukum, atau kepatuhan yang berkelanjutan dari suatu sistem penyimpanan data. Proses ini bukan sekadar penghapusan berkas secara teknis, melainkan melibatkan pertimbangan yang matang terhadap implikasi bisnis, risiko, dan persyaratan regulasi yang terkait dengan berbagai jenis data. Eliminasi data yang efektif memerlukan kebijakan retensi data yang jelas, yang mendefinisikan jangka waktu penyimpanan untuk berbagai kategori informasi berdasarkan nilai bisnis, persyaratan hukum, dan kebutuhan operasional. Proses ini juga melibatkan identifikasi data yang memenuhi kriteria penghapusan, seringkali melalui analisis metadata, usia data, pola akses, dan keterkaitannya dengan proses bisnis aktif. Verifikasi yang cermat diperlukan untuk memastikan bahwa data yang akan dihapus memang tidak lagi dibutuhkan dan tidak akan berdampak negatif pada operasi atau analisis di masa depan. Menurut Erhard dan Hong (2000), proses eliminasi data mencakup kegiatan analisis metadata, usia data, dan tingkat keterpakaian informasi terhadap proses bisnis aktif. Dalam konteks big data dan sistem informasi modern, eliminasi data juga mencakup penggunaan algoritma dan logika pemrograman, seperti SQL, untuk mendeteksi dan menghapus entri data yang bersifat redundan.

Normalisasi Database

Normalisasi database diperkenalkan oleh Edgar F. Codd sebagai prinsip dalam desain struktur basis data relasional. Normalisasi bertujuan untuk menghilangkan redundansi dan memastikan integritas data melalui pembentukan tabel yang saling berkaitan secara logis. Menurut Connolly dan Begg (2015), proses normalisasi terdiri dari beberapa bentuk normal (normal forms) yang disusun untuk meminimalkan duplikasi dan mengoptimalkan struktur penyimpanan data. Dalam konteks jurnal ini, eliminasi data melalui SQL merupakan langkah lanjut dari prinsip normalisasi, terutama ketika data yang sudah terlanjur redundant tetap harus dibersihkan secara teknis. Meskipun

normalisasi mencegah duplikasi sejak awal, metode eliminasi berperan sebagai koreksi sistem terhadap kegagalan desain atau proses entri yang tidak terkendali. Dengan demikian, teori ini memperkuat argumentasi bahwa penanganan duplikasi data bukan hanya soal teknis, tetapi juga berkaitan dengan prinsip desain basis data yang baik.

METODE

Metode penelitian dalam mengatasi duplikasi data dalam database menggunakan metode eliminasi data dilakukan dengan tahapan sistematis yang berfokus pada identifikasi, pemrosesan, dan penghapusan data yang bersifat rangkap atau duplikat untuk meningkatkan kualitas database. Proses penelitian dimulai dengan pengumpulan data dari database yang menjadi objek penelitian, kemudian dilakukan analisis terhadap data tersebut untuk mendeteksi adanya duplikasi berdasarkan atribut-atribut tertentu yang menjadi indikator kemiripan data. Setelah identifikasi dilakukan, metode eliminasi data diterapkan dengan cara membandingkan setiap data yang ada untuk menemukan data yang memiliki kesamaan atau representasi rangkap dari entitas yang sama. Data duplikat yang ditemukan kemudian dihapus sehingga hanya menyisakan satu data yang valid dan unik dalam database. Proses ini bertujuan untuk mengurangi redundansi data dan meningkatkan efisiensi penyimpanan serta keakuratan informasi yang tersimpan. Selain itu, penelitian ini juga melibatkan penggunaan perangkat lunak atau tools khusus yang mendukung proses eliminasi data, seperti fitur penghapusan duplikat pada aplikasi pengolahan data, yang memungkinkan pelaksanaan metode ini secara otomatis dan efektif. Tahapan penelitian ini juga mencakup evaluasi hasil eliminasi untuk memastikan bahwa data yang tersisa sudah bersih dari duplikasi dan siap digunakan untuk keperluan lebih lanjut. Dengan demikian, metode eliminasi data dalam penelitian ini berperan penting dalam menjaga integritas dan kualitas database dengan menghilangkan data yang tidak diperlukan tanpa mengaitkannya pada variabel lain.Selain itu penelitian ini juga mengambil data dari penelitian terdahulu yang dikembangkan agar adanya informasi terbaru tentang eliminasi data.

HASIL DAN PEMBAHASAN

Database, dalam lanskap teknologi informasi kontemporer, memegang peran sentral sebagai fondasi utama dalam pengelolaan informasi. Lebih dari sekadar sekumpulan data yang tersimpan secara digital, database merupakan sistem yang terorganisir secara logis dan terstruktur, dirancang untuk menyimpan, mengelola, dan mengambil data secara efisien dan efektif. Evolusi database dari sistem berkas sederhana hingga platform yang kompleks dan terdistribusi mencerminkan pertumbuhan eksponensial data dan meningkatnya kebutuhan akan pengelolaan informasi yang canggih.Inti dari sebuah database adalah struktur data yang terdefinisi dengan baik. Model data yang beragam, mulai dari model relasional yang klasik dengan tabel-tabel yang saling berhubungan,

hingga model NoSQL yang lebih fleksibel untuk data tidak terstruktur atau semi-terstruktur, memungkinkan organisasi untuk merepresentasikan berbagai jenis informasi sesuai dengan kebutuhan spesifik mereka. Struktur ini memungkinkan data untuk diakses, dimanipulasi, dan dianalisis secara sistematis, membuka pintu bagi berbagai aplikasi dan layanan yang kita gunakan sehari-hari.Pengelolaan database tidak terlepas dari *Database Management System (DBMS)*. Perangkat lunak ini bertindak sebagai perantara krusial antara pengguna, aplikasi, dan data fisik yang tersimpan. DBMS menyediakan serangkaian fungsi vital, termasuk definisi skema database, manipulasi data melalui bahasa kueri seperti SQL, kontrol akses dan keamanan data, pemeliharaan integritas dan konsistensi data, serta mekanisme backup dan pemulihan. Keberadaan DBMS memungkinkan banyak pengguna dan aplikasi untuk berinteraksi dengan data secara bersamaan tanpa mengorbankan integritas dan keamanan.

Signifikansi database meluas ke hampir seluruh aspek kehidupan modern. Dalam dunia bisnis, database menjadi tulang punggung untuk mengelola inventaris, transaksi pelanggan, data keuangan, sumber daya manusia, dan rantai pasokan. Dalam ilmu pengetahuan dan penelitian, database menyimpan dan mengorganisir data eksperimen, hasil observasi, dan literatur ilmiah. Dalam sektor publik, database mendukung layanan pemerintahan, catatan kependudukan, dan sistem informasi publik lainnya. Bahkan dalam interaksi sehari-hari, kita tanpa sadar berinteraksi dengan database saat menggunakan media sosial, berbelanja daring, atau mengakses layanan perbankan.Namun, pengelolaan database juga menghadirkan berbagai tantangan. Pertumbuhan volume data yang eksponensial (*big data*) menuntut solusi penyimpanan dan pemrosesan yang skalabel dan efisien. Kompleksitas model data dan kebutuhan integrasi dengan berbagai sistem yang berbeda memerlukan keahlian khusus dalam desain dan administrasi database. Selain itu, isu-isu seperti keamanan data, privasi, kinerja, dan ketersediaan tinggi menjadi perhatian utama dalam pengelolaan database modern.Hal ini di dukung oleh hasil penelitian yang dilakukan oleh (Ade 2024) yang dimana dalam jurnal yang dikaji,hasil yang didapatkan adalah berbagai celah dalam system jaminan sosial ketenagakerjaan menyebabkan kerentanan dalam keamanan data dapat muncuk dalam tahap pengumpulan data peserta dan *stake holder*.

Dalam konteks pengelolaan database dan mengatasi duplikasi data, eliminasi data memainkan peran ganda. Setelah data duplikat berhasil diidentifikasi, eliminasi menjadi langkah kunci untuk menghapus salinan redundan, menciptakan sumber informasi yang tunggal dan terpercaya. Selain itu, eliminasi data secara berkala membantu memelihara kesehatan dan kinerja database dengan mengurangi volume data yang harus dikelola dan diproses. Dengan demikian, eliminasi data bukan hanya praktik kebersihan digital, tetapi merupakan strategi proaktif untuk mengelola risiko, mengoptimalkan sumber daya, memastikan kepatuhan, dan meningkatkan nilai informasi organisasi secara

keseluruhan. Implementasi proses eliminasi data yang efektif membutuhkan pemahaman yang mendalam tentang data organisasi, kebijakan yang jelas, prosedur yang aman, dan kesadaran akan pentingnya pengelolaan informasi yang bertanggung jawab. Berfokus pada pentingnya pengelolaan data yang bersih dan efisien untuk menjaga kualitas serta integritas database. Duplikasi data merupakan masalah umum yang dapat menyebabkan inkonsistensi informasi, pemborosan ruang penyimpanan, dan menurunkan performa sistem database. Oleh karena itu, metode eliminasi data menjadi solusi utama yang efektif untuk mengidentifikasi dan menghapus data yang berulang sehingga database menjadi lebih terorganisir dan efisien. (Jonathan,Felix 2024)

SQL sendiri merupakan salah satu program yang dipakai dalam metode eliminasi data. Sistem SQL (*Structured Query Language*) memiliki peran penting dalam proses eliminasi data duplikat dalam sebuah database. SQL menyediakan berbagai perintah dan fungsi yang memungkinkan pengguna untuk mengidentifikasi, memfilter, dan menghapus data yang berulang secara efisien. Salah satu metode yang umum digunakan adalah dengan memanfaatkan fungsi ROW_NUMBER() yang dipadukan dengan klausa PARTITION BY, yang berfungsi untuk memberi nomor urut pada data berdasarkan kelompok tertentu, sehingga baris duplikat dapat dikenali dan dihapus dengan mudah. Selain itu, perintah DISTINCT juga dapat digunakan untuk menampilkan hanya data yang unik dari suatu tabel. Dalam konteks manajemen data perusahaan, seperti Indofood yang menangani ribuan data produk, pelanggan, dan transaksi, penerapan sistem SQL untuk eliminasi data sangat krusial dalam menjaga akurasi dan integritas informasi. Dengan demikian, SQL tidak hanya membantu dalam pengelolaan basis data, tetapi juga mendukung pengambilan keputusan yang lebih tepat berbasis data yang bersih dan terstruktur. (Didik & Herlawati 2019)

Berikut contoh data yang sudah dan belum di eliminasi menggunakan program eliminasi data (SQL) dengan rumus:

```
SELECT nama, email, COUNT(*)
FROM pelanggan
GROUP BY nama, email
HAVING COUNT(*) > 1;
```

Tabel 1

Id Pelanggan	Nama	Email
1	Andi	andi@gmail.com
2	Budi	budi@gmail.com
3	Andi	andi@gmail.com
4	Sari	sari@gmail.com
5	Budi	budi@gmail.com

Tabel 2

Count	Nama	Email
2	Andi	andi@gmail.com
2	Budi	budi@gmail.com
1	Sari	sari@gmail.com

Jika diihat saja secara sederhana program SQL sangat memudahkan dalam penginputan dan data yang didapatkan.Tidak adanya pemborosan ruang dan juga pencarian data yang masuk lebih trestruktur.Hal ini berkenan dengan meningkatkan kualitas data dan menghindari duplikasi data.Diera modern duplikasi data merupakan hal yang bersinggungan langsung dengan konerja perusahaan.Pengambilan keputusan juga menimbang akan kualitas data,maka hasil akhir dari pengambilan keputusan akan sangat berpengaruh terhadap kualitas data.

Mengatasi duplikasi data dalam database menggunakan metode eliminasi data merupakan langkah strategis yang sangat penting dalam menjaga kualitas, konsistensi, dan integritas informasi yang tersimpan. Duplikasi data tidak hanya menimbulkan masalah teknis seperti pemborosan ruang penyimpanan dan penurunan performa sistem, tetapi juga berdampak negatif pada keakuratan analisis dan pengambilan keputusan yang bergantung pada data tersebut. Dengan adanya data yang berulang, organisasi berisiko menghasilkan laporan yang tidak valid dan kebijakan yang kurang tepat, sehingga dapat merugikan operasional dan reputasi perusahaan. Oleh karena itu, penerapan metode eliminasi data menjadi solusi efektif untuk mengidentifikasi dan menghapus data duplikat secara sistematis. (Erhard & Hong 2000)

Metode eliminasi data, khususnya yang diimplementasikan melalui bahasa SQL, menawarkan berbagai teknik yang memungkinkan proses deduplikasi berjalan otomatis dan efisien. Dengan demikian, metode ini tidak hanya membersihkan database dari data yang tidak perlu, tetapi juga meningkatkan kecepatan akses dan pengolahan data, serta mengoptimalkan penggunaan sumber daya penyimpanan.Lebih jauh, eliminasi data juga berperan penting dalam menjaga integritas dan keandalan sistem informasi secara menyeluruh. Data yang bersih dan terstruktur rapi memudahkan proses audit, pemeliharaan, dan pengembangan database di masa depan. Hal ini juga mendukung terciptanya ekosistem data yang sehat, di mana setiap informasi yang diolah dan disajikan dapat dipertanggungjawabkan secara akurat. Oleh karena itu, penguasaan dan penerapan metode eliminasi data menjadi keterampilan yang esensial bagi para pengelola database dan pengembang sistem informasi. Dengan memanfaatkan teknik SQL dan alat bantu lainnya secara optimal, organisasi dapat mengatasi tantangan duplikasi data secara efektif, sehingga mendukung pengambilan keputusan yang lebih tepat, efisien, dan berkelanjutan. Secara keseluruhan, penerapan metode eliminasi data adalah fondasi penting

dalam membangun sistem database yang handal dan berkualitas tinggi, yang pada akhirnya meningkatkan daya saing dan keberhasilan organisasi di era digital saat ini. (Ghassani,A., Setiaji, H 2019)

KESIMPULAN DAN SARAN

Dalam era digital yang terus berkembang, pengelolaan data menjadi pondasi penting bagi berbagai sektor, termasuk bisnis, pemerintahan, kesehatan, dan pendidikan. Database memainkan peran sentral dalam menyimpan, mengelola, dan mengambil data yang terstruktur. Namun, salah satu tantangan utama dalam pengelolaan data adalah munculnya duplikasi data, yang dapat mengakibatkan berbagai masalah serius. Duplikasi data, yaitu keberadaan salinan data yang sama atau sangat mirip dalam database, dapat disebabkan oleh kesalahan input manual, integrasi data dari berbagai sumber yang kurang baik, atau kurangnya mekanisme validasi yang efektif. Dampaknya meliputi inkonsistensi informasi, pemborosan ruang penyimpanan, ketidakakuratan analisis, dan pengambilan keputusan yang salah.

Untuk mengatasi masalah ini, metode eliminasi data menjadi sangat penting. Metode ini melibatkan identifikasi dan penghapusan data duplikat untuk menjaga kualitas, konsistensi, dan integritas database. Salah satu alat yang efektif dalam implementasi eliminasi data adalah SQL (Structured Query Language). SQL memungkinkan pengelompokan data, identifikasi data ganda, dan penghapusan baris yang berulang, sehingga menghasilkan database yang lebih bersih dan terorganisir.Penerapan metode eliminasi data tidak hanya memperbaiki masalah teknis tetapi juga memiliki manfaat strategis. Data yang bersih meningkatkan akurasi laporan, mendukung pengambilan keputusan yang lebih tepat, mengoptimalkan penggunaan sumber daya penyimpanan, dan meningkatkan kinerja sistem. Selain itu, eliminasi data mempermudah audit, pemeliharaan, dan pengembangan database di masa depan, yang sangat penting dalam ekosistem data yang sehat.

Sebagai langkah proaktif dalam mengatasi duplikasi data, organisasi harus menerapkan strategi pencegahan yang kuat. Ini mencakup penerapan mekanisme validasi data yang ketat pada titik masuk data, seperti formulir daring dan aplikasi entri data, untuk meminimalkan risiko kesalahan input manual yang merupakan salah satu penyebab utama duplikasi. Teknologi deduplikasi data waktu nyata dapat digunakan untuk mengidentifikasi dan mencegah duplikasi saat data dimasukkan ke dalam sistem, memberikan lapisan perlindungan tambahan. Selain itu, audit data rutin harus dilakukan untuk mendeteksi dan membersihkan duplikasi yang mungkin telah terjadi, memastikan bahwa database tetap bersih dan akurat.Dalam konteks eliminasi data, SQL (Structured Query Language) memainkan peran penting sebagai alat yang ampuh untuk manipulasi data. Organisasi harus memastikan bahwa tim pengelola basis data mereka memiliki keahlian mendalam dalam SQL dan mampu memanfaatkan

fitur-fiturnya secara efektif untuk mengidentifikasi, memfilter, dan menghapus data duplikat.

REFERENSI

Ade H. (2024) "Mengintip Celah antara Potensi dan Tantangan Big Data pada Layanan Jaminan Sosial Ketenagakerjaan Indonesia". Jurnal Jamsostek 2.2 .Hal.194-196.

Ainayya G, Hari S. (2019)"Data Cleansing Pada Data Rumah Sakit."proceeding SINTAK : 242-248.

Connolly, T., & Begg, C. (2015). Database systems: A practical approach to design, implementation, and management (6th ed.). Pearson Education.

Didik, S , Herlawati. (2019). "Structured Query Language (SQL) Untuk Purchase Order (PO) Menggunakan SQL Server 2008". Bina Insani Ict Journal 6.1.Hal.76-80.

Erhard, R , Hong, H. (2000),"Data Cleaning: Problems and Current Approaches". Hal. 9-10.

Hery Fadly, & Deddy Yusuf. (2024). Teknologi Media Digital Terhadap Produktifitas Serta Dampanya Pada Bidang bisnis, dan Sosial budaya (Studi Kasus Desa Bone Kec. Bajeng Kab. Gowa Sulawesi Selatan). *JUMABI: Jurnal Manajemen, Akuntansi Dan Bisnis*, *2*(1), 21–28. https://doi.org/10.56314/jumabi.v2i1.193

Intan F, Irwan M. (2024), "Pengenalan Sistem Database : Konsep Dasar Dan Manfaatnya Dalam Perusahaan". Kampus Akademik Publising Jurnal Ilmiah Nusantara (JINU) 1.4 .Hal. 673-678.

Jonathan, K, W, Felix, R, Et Al. (2024), "Data Mining (Memahami Pola di Balik Data)". Eureka Media Aksara.Hal. 38-46.

K, S., Suryaningsih, S., & Irsang, I. (2025). Analisis Pengaruh Digital Marketing Terhadap Peningkatan Penjualan Hasil Industri Rumahan Desa Samataring Kec. Kelara Kab. Jeneponto. *JUMABI: Jurnal Manajemen, Akuntansi Dan Bisnis*, *3*(1), 11–19. https://doi.org/10.56314/jumabi.v3i1.279

"Manajemen Proyek Informasi" (2012):. Uinsu Buku Pengajar, Tantra.Hal. 91-101.

Rismaninda P, Radifa. (2025 "Implementasi Penggunaan Data Analytics untuk Mengoptimalkan Pengambilan Keputusan Bisnis di Era Digital." Jurnal Bisnis dan Komunikasi Digital 2.2: Hal. 1-12.

Susilo H, Yudho G. (2020), "Pengukuran Kualitas Data Menggunakan Framework Total Data Quality Management (TDQM): Studi Kasus Sistem Informasi Beasiswa Universitas Indonesia". Jurnal IPTEK-KOM 22.2 .Hal.193-206.

Wali, M., Efitra. (2023) "Penerapan Dan Implementasi Big Data di Berbagai Sektor". Son Pedia Publishing Indonesia Hal.50-61.