

Optimalisasi Database Mysql Pada Sistem Perancangan Sistem Housekeeping Transaction History Dengan Pentaho Data Integration

Yohanes Albryan Simanjuntak¹, Alz Danny Wowor²

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Indonesia

E-mail : 672020151@student.uksw.edu¹, alzdanny.wowor@uksw.edu²

Abstract

The volume of transaction data in the banking industry is growing with the increase in customers and transaction complexity. Inefficient data management can lead to server overload, affect system performance, and hinder the delivery of fast, accurate services. Housekeeping processes are needed to move inactive data to separate storage, allowing the main server to function more efficiently. Pentaho Data Integration (PDI) offers an effective solution for managing the ETL (Extract, Transform, Load) process, which is crucial for data housekeeping. This research aims to optimize the management of banking transaction data using PDI to reduce server load and improve operational efficiency. This quantitative study applies an experimental method, with the ETL process managing Bank XYZ's transaction data older than six months. The study uses transaction data from Bank XYZ's MySQL server, which will be transferred to a data warehouse. The analysis applies clustering algorithms to filter and separate active from inactive transactions. The implementation of PDI for housekeeping effectively reduces server load and improves data management efficiency, significantly lowering processing time. The combined use of clustering algorithms and PDI delivers substantial improvements in managing banking transaction data, enhancing operational efficiency while significantly reducing the load on the main server.

Keywords: Clustering Algorithms, Data Management Efficiency, ETL Proses, Pentaho Data Integration, Transaction Data Management

Abstrak

Volume data transaksi di industri perbankan terus meningkat seiring dengan pertumbuhan jumlah nasabah dan kompleksitas transaksi. Pengelolaan data yang tidak efisien dapat menyebabkan beban server yang berlebih, mempengaruhi kinerja sistem, dan mengakibatkan kendala dalam penyediaan layanan yang cepat dan akurat. Proses housekeeping diperlukan untuk memindahkan data yang sudah tidak aktif ke penyimpanan terpisah agar server utama dapat berfungsi lebih optimal. Pentaho Data Integration (PDI) menyediakan solusi yang efisien untuk mengelola proses ETL (Extract, Transform, Load), yang penting dalam proses housekeeping data transaksi. Penelitian ini bertujuan untuk mengoptimalkan pengelolaan data transaksi perbankan melalui pendekatan housekeeping dengan menggunakan Pentaho Data Integration, sekaligus mengurangi beban pada server utama dan meningkatkan efisiensi operasional. Jenis penelitian ini adalah penelitian kuantitatif yang memanfaatkan metode eksperimen. Proses ETL diterapkan untuk mengelola data transaksi dari Bank XYZ yang berusia lebih dari 6 bulan. Subjek penelitian adalah data transaksi dari server MySQL Bank XYZ, yang akan dipindahkan ke data warehouse. Teknik analisis data melibatkan penerapan algoritma clustering untuk menyaring data dan memisahkan transaksi aktif dari yang sudah lama. Implementasi Pentaho Data Integration untuk proses housekeeping terbukti efektif dalam mengurangi beban server dan meningkatkan efisiensi pengelolaan data transaksi, di mana proses ETL yang diterapkan berhasil menurunkan waktu pemrosesan secara signifikan. Proses house-keeping yang dilakukan dengan bantuan algoritma

clustering dan PDI memberikan peningkatan kinerja yang signifikan dalam pengelolaan data transaksi perbankan. Efisiensi operasional meningkat, dan beban pada server utama berkurang secara substansial.

Kata kunci: *Algoritma Clustering, Efisiensi Operasional, ETL (Extract, Transform, Load), Pentaho Data Integration, Transaksi Perbankan*

1. Pendahuluan

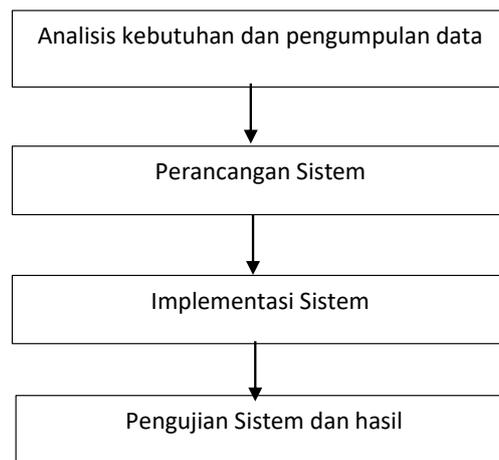
Industri perbankan telah mengalami pertumbuhan pesat, baik dalam hal jumlah nasabah maupun volume transaksi. Dalam beberapa tahun terakhir, teknologi informasi dan komunikasi memainkan peran penting dalam memfasilitasi transaksi keuangan yang cepat dan aman. Namun, seiring dengan peningkatan volume data transaksi, bank menghadapi tantangan besar dalam mengelola dan menyimpan data secara efisien. Sistem transaksi perbankan harus tetap handal dan memenuhi standar regulasi, seperti pelaporan audit dan penyimpanan data dalam jangka waktu tertentu [1]. Masalah utama yang dihadapi oleh Bank XYZ adalah peningkatan volume data transaksi yang secara signifikan membebani infrastruktur TI mereka. Hal ini menyebabkan waktu akses data yang lambat, meningkatnya waktu respons, dan kesulitan dalam menjalankan analisis data secara cepat. Selain itu, penyimpanan dan pengelolaan data yang tidak efisien juga berdampak pada kemampuan bank untuk memenuhi kebutuhan audit dan kepatuhan terhadap regulasi. Untuk mengatasi permasalahan ini, diperlukan pendekatan housekeeping yang dapat memindahkan data transaksi lama ke data warehouse dengan menggunakan Pentaho Data Integration (PDI). Pentaho merupakan nama suatu perusahaan dan juga nama produk. Pentaho mengkhususkan diri sebagai perusahaan pembuat software untuk keperluan data warehouse dan *business intelligence* (BI) [2]. Pentaho adalah sebuah software yang digunakan dalam proses ETL dalam penyusunan data warehouse [3]. ETL adalah sekumpulan proses untuk mengambil dan memproses data dari satu atau banyak sumber data menjadi sumber baru [4]. Tools ETL mengekstrak data dari sumber yang dipilih, mengubahnya menjadi format baru sesuai dengan aturan bisnis, dan kemudian memuatnya ke dalam struktur data target [5]. ETL merupakan proses yang sangat penting dalam data warehouse, dengan ETL inilah data dari operasional dapat dimasukkan ke dalam data warehouse [6]. Elemen utama dari *Pentaho Data Integration* (PDI) adalah *Transformation* dan *Job*. *Transformation* adalah sekumpulan instruksi untuk merubah input menjadi output yang diinginkan (input-proses-output). Sedangkan *Job* adalah kumpulan instruksi untuk menjalankan transformasi [7]. Pendekatan ini melibatkan proses ETL untuk menyaring data yang sudah tidak aktif dan mengoptimalkan penyimpanan serta akses data yang lebih efisien. Housekeeping merupakan bagian penting dari manajemen database, yang bertujuan untuk mengoptimalkan kinerja sistem dengan memisahkan data yang aktif dan tidak aktif.

Berdasarkan penelitian [1], proses ETL dengan Pentaho Data Integration telah terbukti meningkatkan efisiensi dalam manajemen data transaksi perbankan. Selain itu, penelitian [8], menunjukkan bahwa struktur database yang baik mampu mendukung analisis data secara lebih cepat dan akurat. Beberapa penelitian sebelumnya telah membahas penggunaan *Pentaho Data Integration* dalam proses ETL, tetapi belum ada yang secara spesifik mengeksplorasi penerapan algoritma clustering untuk optimalisasi housekeeping pada data transaksi perbankan. Penelitian ini diharapkan dapat mengisi gap tersebut dengan menawarkan solusi berbasis clustering untuk proses housekeeping. Clustering adalah proses pengelompokan kumpulan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek dikelompok lain [9]. Teknologi clustering data merupakan suatu teknik yang menunjukkan persamaan karakteristik dalam suatu kelompok sehingga akan menghasilkan informasi yang bermanfaat [10]. Algoritma clustering data

sudah banyak dipergunakan diberbagai bidang [11]. Oleh karena itu, clustering sangat berguna dan bisa menemukan grup yang tidak dikenal dalam data. Cluster analysis banyak digunakan dalam berbagai aplikasi seperti *Business Intelligence*, *Image Pattern Recognition*, *Web Search*, *Biology*, dan *Security* [12]. Di dalam *business intelligence*, clustering bisa mengatur banyak *customer* ke dalam banyak grup. *Clustering* juga dikenal sebagai data segmentation, karena clustering mempartisi banyak data set ke dalam banyak grup berdasarkan persamaannya. *Clustering* juga bisa sebagai *outlier detection*, di mana *outlier* bisa menjadi menarik daripada kasus yang biasa [13]. Pendekatan ini bertujuan untuk mengurangi beban infrastruktur server dengan melakukan penyaringan data transaksi yang tidak aktif, memudahkan proses penyimpanan dan akses data transaksi untuk kebutuhan audit dan kepatuhan regulasi dan meningkatkan efisiensi operasional bank dalam mengelola data transaksi secara keseluruhan. Dengan mencapai tujuan ini, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan bagi industri perbankan dalam menghadapi tantangan pengelolaan data transaksi yang semakin kompleks. Diharapkan, hasil penelitian ini dapat memberikan solusi praktis untuk mengurangi beban server dan meningkatkan efisiensi operasional sistem transaksi di Bank XYZ.

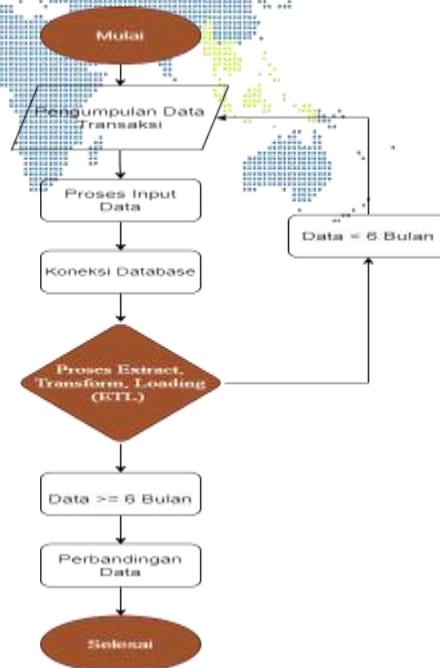
2. Metodologi Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan metode eksperimen. Metode ini melibatkan perancangan, implementasi, dan pengujian sistem housekeeping yang menggunakan algoritma clustering untuk mengoptimalkan pengelolaan data transaksi di server MySQL. Desain penelitian meliputi beberapa tahap, yaitu analisis kebutuhan, pengumpulan data, perancangan sistem, implementasi sistem, dan pengujian sistem.



Gambar 1. Diagram Alir Penelitian

Pada tahap awal, data transaksi dari server Bank XYZ dianalisis untuk menentukan volume data yang perlu dipindahkan. Setelah itu, sistem dirancang untuk menerapkan proses ETL menggunakan Pentaho Data Integration, yang berfungsi untuk memindahkan data transaksi yang berusia lebih dari 6 bulan ke dalam data warehouse.



Gambar 2. Proses Penelitian

Subjek penelitian adalah data transaksi perbankan dari Bank XYZ yang sudah tidak aktif (lebih dari 6 bulan). Data ini diambil dari server MySQL Bank XYZ dan diproses melalui Pentaho Data Integration sebelum dipindahkan ke data warehouse. Pengumpulan data dilakukan dengan menggunakan data dummy untuk mensimulasikan kondisi transaksi nyata. Data tersebut diolah melalui proses ETL (Extract, Transform, Load), di mana data transaksi lama dipisahkan dari data transaksi aktif. Setiap transaksi diklasifikasikan berdasarkan tanggal transaksi untuk menentukan kelayakannya dalam proses housekeeping. Data dianalisis menggunakan algoritma clustering untuk menyaring dan mengelompokkan transaksi berdasarkan masa aktifnya. Transaksi yang berusia lebih dari 6 bulan dipindahkan ke data warehouse untuk mengurangi beban server utama. Waktu pemrosesan sebelum dan sesudah implementasi dibandingkan untuk menilai efisiensi sistem.

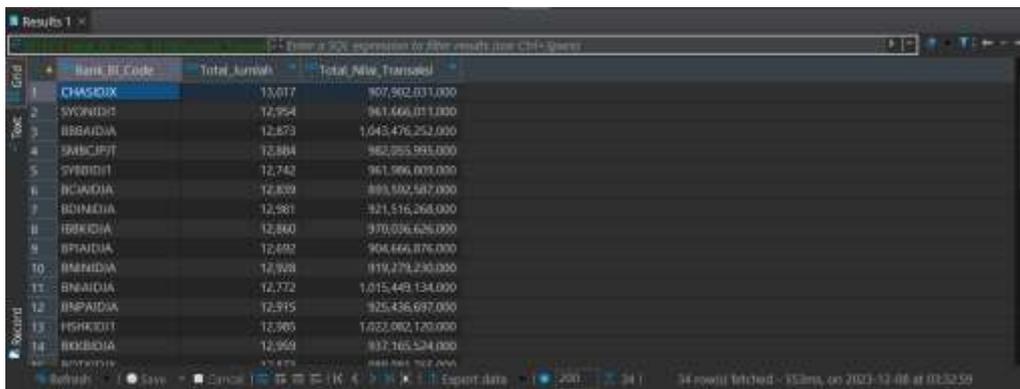
3. Hasil dan Pembahasan

Implementasi sistem housekeeping menggunakan *Pentaho Data Integration* menunjukkan hasil yang signifikan dalam mengurangi beban server. Data transaksi lama yang telah dipindahkan ke data warehouse berhasil menurunkan waktu akses data di server utama. Berdasarkan pengujian, waktu pemrosesan data sebelum housekeeping mencapai 553ms, sementara setelah proses housekeeping dilakukan, waktu pemrosesan turun menjadi 1.137s (2ms). Selain itu, hasil analisis menunjukkan bahwa proses clustering membantu dalam mengelompokkan transaksi lama dengan lebih efisien, sehingga hanya data aktif yang diproses oleh server utama. Pengelompokkan ini mempunyai tujuan mengelompokkan objek menurut karakteristik antar objek. Dari analisis cluster, kita dapat menemukan grup yang memiliki karakteristik masing-masing grup [14].

```

SELECT Bank_BI_Code, SUM(Jumlah_Transaksi) AS Total_Jumlah, SUM(nilai_transaksi) AS
FROM (
SELECT bank_bi_code AS Bank_BI_Code,
COUNT(*) AS Jumlah_Transaksi,
nilai_transaksi
FROM history_transaksi_bifast
GROUP BY bank_bi_code, nilai_transaksi
UNION ALL
SELECT bank_bi_code AS Bank_BI_Code,
COUNT(*) AS Jumlah_Transaksi,
nilai_transaksi
FROM history_transaksi_realtime
GROUP BY bank_bi_code, nilai_transaksi
UNION ALL
SELECT bank_bi_code AS Bank_BI_Code,
COUNT(*) AS Jumlah_Transaksi,
nilai_transaksi
FROM history_transaksi_rtgs
GROUP BY bank_bi_code, nilai_transaksi
UNION ALL
SELECT bank_bi_code AS Bank_BI_Code,
COUNT(*) AS Jumlah_Transaksi,
nilai_transaksi
FROM history_transaksi_skn
GROUP BY bank_bi_code, nilai_transaksi
UNION ALL
SELECT bank_bi_code AS Bank_BI_Code,
COUNT(*) AS Jumlah_Transaksi,
nilai_transaksi
FROM history_transaksi_omax
GROUP BY bank_bi_code, nilai_transaksi
) AS subquery
GROUP BY Bank_BI_Code;
    
```

Gambar 3. Query sql untuk melakukan testing data

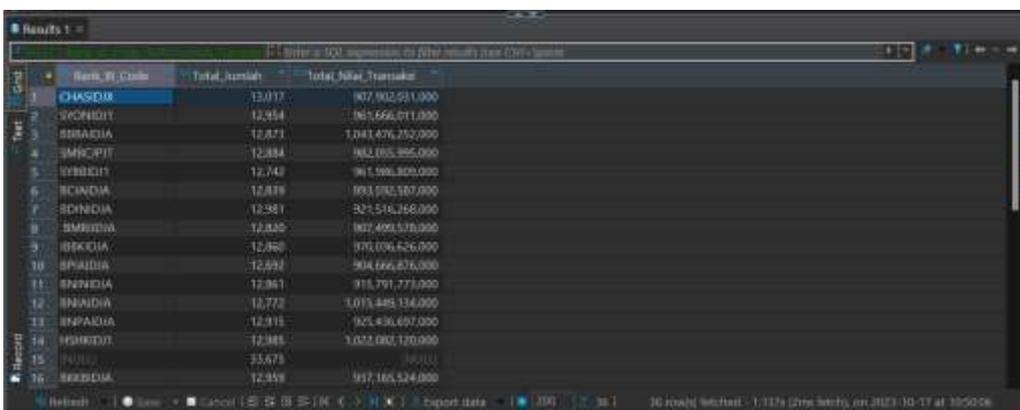


Bank_BI_Code	Total_Jumlah	Total_Nilai_Transaksi
CHASDIK	13,017	907,902,031,000
SYONDIK	12,954	961,666,011,000
BRBAIDIA	12,873	1,043,476,252,000
SMKCPIT	12,884	982,255,995,000
SYBIDIK	12,742	961,996,800,000
BCWADIA	12,809	889,582,587,000
BDMADIA	12,981	321,516,268,000
BRKADIA	12,860	370,036,626,000
BPAIDIA	12,692	904,666,876,000
BMAIDIA	12,968	819,279,230,000
BNAIDIA	12,772	1,015,448,134,000
BMPADIA	12,915	325,436,697,000
HSHKDIK	12,985	1,032,082,120,000
BKKBADIA	12,959	837,165,524,000

Gambar 4. Tabel output dari hasil query

34 34 row(s) fetched - 553ms, 0

Gambar 5. Hasil waktu yang dibutuhkan



Bank_BI_Code	Total_Jumlah	Total_Nilai_Transaksi
CHASDIK	13,017	907,902,031,000
SYONDIK	12,954	961,666,011,000
BRBAIDIA	12,873	1,043,476,252,000
SMKCPIT	12,884	982,255,995,000
SYBIDIK	12,742	961,996,800,000
BCWADIA	12,809	889,582,587,000
BDMADIA	12,981	321,516,268,000
BRKADIA	12,860	370,036,626,000
BPAIDIA	12,692	904,666,876,000
BMAIDIA	12,968	819,279,230,000
BNAIDIA	12,772	1,015,448,134,000
BMPADIA	12,915	325,436,697,000
HSHKDIK	12,985	1,032,082,120,000
BKKBADIA	12,959	837,165,524,000

Gambar 6. Tabel output dari hasil query

36 row(s) fetched - 1.137s (2ms fetch), on 2023-10-17

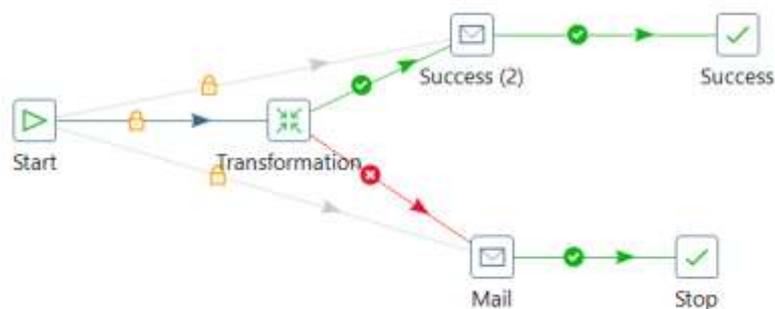
Gambar 7. Hasil waktu yang dibutuhkan

Hasil ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa proses ETL, terutama dengan Pentaho Data Integration, dapat mengoptimalkan kinerja sistem manajemen data transaksi perbankan [15]. Dalam hal ini, *Extract Transform Load* (ETL) process akan menjadi metodologi utama untuk mengintegrasikan data dari berbagai sumber dan juga data transforming pada data yang terintegrasi [16]. ETL mencakup tiga langkah yaitu ekstraksi, transformasi, dan pengiriman data dari berbagai sumber ke satu penyimpanan sentral agar akses data yang diperoleh lebih akurat dan terstruktur [17]. Implementasi algoritma clustering dalam proses housekeeping memberikan keuntungan tambahan dalam hal efisiensi penyimpanan dan pemrosesan data. Penggunaan clustering memungkinkan pemisahan data secara lebih akurat dan cepat, yang tidak hanya meningkatkan kecepatan akses data tetapi juga mengurangi redundansi [18].

Temuan penelitian ini mendukung teori manajemen database yang menekankan pentingnya housekeeping dalam menjaga integritas dan keandalan sistem. Dengan adanya sistem keamanan yang tepat, akses ke data sensitif dapat dikendalikan dan informasi rahasia dapat dilindungi dari akses yang tidak sah. Integritas data juga dijaga dengan adanya aturan validasi dan pengendalian yang diterapkan pada database, mencegah kesalahan atau manipulasi data yang tidak sah [8]. Secara keseluruhan, database memiliki pengaruh yang signifikan terhadap Sistem Informasi Manajemen. Dengan menyediakan akses yang cepat, integrasi data, keamanan, dan analisis yang lebih baik, database yang baik dapat meningkatkan efisiensi, efektivitas, dan kinerja manajemen organisasi. Oleh karena itu, penting bagi organisasi untuk mengelola dan memelihara database dengan baik guna mendukung kesuksesan sistem informasi manajemen mereka. [19]. Namun, perbedaan utama dari penelitian sebelumnya adalah penggunaan algoritma clustering yang menawarkan fleksibilitas dan efisiensi yang lebih besar dalam mengelola data transaksi yang besar dan kompleks [20].

4. Kesimpulan

Penelitian ini membuktikan bahwa penerapan algoritma clustering dalam proses housekeeping dengan menggunakan Pentaho Data Integration dapat secara signifikan meningkatkan efisiensi operasional dan mengurangi beban server pada sistem transaksi perbankan. Dengan memindahkan data transaksi yang tidak aktif ke data warehouse, sistem dapat mengelola transaksi secara lebih efisien dan cepat. Selain itu, implementasi ini memberikan kontribusi penting dalam teori manajemen database dengan menawarkan pendekatan baru dalam proses housekeeping data transaksi.



Gambar 8. Proses ETL dengan Pentaho Data Integration

Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi lebih lanjut analisis biaya-manfaat dari implementasi ini, serta dampak dari faktor organisasi lain yang mungkin mempengaruhi hasil optimalisasi data transaksi di lingkungan perbankan.

Daftar Pustaka

- [1] I. P. W. Prasetya and I. N. H. Kurniawan, "Implementasi ETL (Extract, Transform, Load) pada Data warehouse Penjualan Menggunakan Tools Pentaho," *TIERS Inf. Technol. J.*, vol. 2, no. 1, pp. 1–8, 2021, doi: 10.38043/tiers.v2i1.2844.
- [2] M. R. Hartani and I. B. M. Mahendra, "Implementasi Data Warehouse Menggunakan Pentaho BI di Hartaning House Homestay," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 10, no. 1, p. 153, 2021, doi: 10.24843/jlk.2021.v10.i01.p18.
- [3] W. P. N. Witjaksono, R. Wahjoe., Wiyogo Mardiyanto., "Perancangan Aplikasi Business Intelligence Pada Sistem Informasi Distribusi PT Pertamina Lubricant Menggunakan Pentaho," *J. Rekayasa Sist. Ind.*, vol. 2, no. 2, pp. 12–18, 2015.
- [4] D. Putri, Pipit M., Gunawan, Syam., Shubuh, "Perancangan Data warehouse dan Penerapannya pada Sistem Informasi Penjualan Menggunakan Tools Pentaho dan Tableau," *J. Inform. Komputasi*, vol. 12, no. 1, pp. 15–21, 2018.
- [5] A. Hidayat, Luky., Permanasari, Adhistya Erna., "Perancangan Data warehouse E-Procurement Pada Instansi Pemerintahan," *Semin. Nas. Inov. Dan Apl. Teknol. Di Ind.*, vol. 3, no. 1, pp. 31–36, 2017.
- [6] S. Darudiat, "Perancangan Data warehouse Penjualan untuk Mendukung Kebutuhan Informasi Eksekutif Cemerlang Skin Care," *Semin. Nas. Inform. 2010*, pp. 350–359, 2010.
- [7] S. N. W. Sumartini and M. N. M. Lisma, "Extract Transform Loading Data Absensi STMIK STIKOM Indonesia Menggunakan Pentaho," *J. MATRIK*, vol. 19, no. 2, pp. 273–281, 2020.
- [8] K. Syahputri, M. Irwan, and P. Nasution, "Peran Database Dalam Sistem Informasi Manajemen," *J. Akunt. Keuang. dan Bisnis*, vol. 1, no. 2, pp. 54–58, 2023, [Online]. Available: <https://jurnal.ittc.web.id/index.php/jakbs/article/view/36>
- [9] P. A. Jusia, "Analisis komparasi pemodelan algoritma decision tree menggunakan metode particle swarm optimization dan metode adaboost untuk prediksi awal penyakit jantung," *Semin. Nas. Sist. Inf.*, pp. 1048–1056, 2018.
- [10] S. S. Lubis and B. Hendrik, "Implementasi Data Mining Pengelompokan Data Penjualan Berdasarkan Pembelian Dengan Menggunakan Algoritma K-Means Pada UD.Martua," *J. Inf. Sysem Educ. Dev.*, vol. 1, no. 3, pp. 36–41, 2023, [Online]. Available: <https://journal.widyakarya.ac.id/index.php/jusiik-widyakarya/article/view/1531%0Ahttps://journal.widyakarya.ac.id/index.php/jusii-k-widyakarya/article/download/1531/1563>
- [11] P. A. Jusia, "Face Recognition Menggunakan Metode Algoritma Viola Jones Dalam Penerapan Computer Vision.," *J. Ilm. Media Process.*, vol. 11, no. 1, pp. 663–675, 2016.
- [12] J. P. Jiawei Han, Micheline Kamber, "Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)," 2011.
- [13] P. Alam Jusia, F. Muhammad Irfan, and S. Dinamika Bangsa Jambi Jl Jend Sudirman Thehok Jambi, "Clustering Data Untuk Rekomendasi Penentuan Jurusan Perguruan Tinggi Menggunakan Metode K-Means," *J. IKRA-ITH Inform.*, vol. 3, no. 3, p. 75, 2019.
- [14] F. Handayani, "Aplikasi Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokan Mahasiswa Berdasarkan Gaya Belajar," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 46–63, 2022, doi: 10.34010/jati.v12i1.6733.
- [15] B. K. Easterita, I. Arwani, and D. E. Ratnawati, "Pengembangan Data Warehouse dan Online Analytical Processing (OLAP) untuk Data Artikel pada Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)," *J. Sist. Informasi, Teknol. Informasi, dan Edukasi Sist. Informasi*, vol. 1, no. 1, pp. 12–22, 2020.
- [16] R. Abdul Ghani and R. Kurniawan, "Implementasi Extract, Transform, Load Process Pada Perancangan Data Warehouse Terkait Kualitas Pendidikan Di

- Kabupaten Serang,” *JATI (Jurnal Mhs. Tek. Inform.,* vol. 8, no. 2, pp. 2083–2090, 2024, doi: 10.36040/jati.v8i2.9081.
- [17] S. C. M. Souibgui, F. Atigui, S. Zammali and S. Ben Yahia, “Data quality in ETL process: A preliminary study,” *Procedia Comput.*, vol. 159, pp. 676–687, 2019.
- [18] J. A. O’Brien and G. M. Marakas, *Management Information Systems*. McGraw-Hill Education, 2016.
- [19] I. E. Gordon, “The HITRAN2020 molecular spectroscopic database,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 227, 2022.
- [20] T. Connolly and C. Begg, *A Practical Approach to Design, Implementation, and Management*. Pearson Education, 2014.