



Indonesian Journal of Machine Learning and Computer Science
Institut Riset dan Publikasi Indonesia (IRPI)



JOURNAL INFO

Journal title	: MALCOM: Indonesian Journal of Machine Learning and Computer Science
Initials	: MALCOM
Accredited	: Sinta 3 Kemendikbud (2025-2030)
Frequency	: 4 issues per Year (January, April, July, and October) Issue Change Info
P-ISSN	: 2797-2313
E-ISSN	: 2775-8575
Publisher	: Institut Riset dan Publikasi Indonesia (IRPI)



Editorial Team

Editor in Chief



Mustakim, S.T., M.Kom.

UIN Sultan Syarif Kasim Riau, Pekanbaru - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)

Editorial Board



Assoc. Prof. Dr. Ir. Sumijan, M.Sc.

Universitas Putra Indonesia "YPTK" Padang, West Sumatera- Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Iwan Adhicandra, S.T., M.Sc., Ph.D.

Universitas Bakrie, Jakarta - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Teddie Darmizal, S.T., M.T., Ph.D.

UIN Sultan Syarif Kasim Riau, Pekanbaru - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Dr. Muhardi, S.Kom., M.Kom.
Universitas Dumai, Dumai - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Edi Ismanto, S.T., M.Kom., Ph.D.
Universitas Muhammadiyah Riau, Pekanbaru - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Dr. Erick Fernando, S.Kom., MSI.
Universitas Multimedia Nusantara, Banten - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Muhamad Irsan, S.T., M.Kom.
Telkom University, Bandung, West Java - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Inggih Permana, S.T., M.Kom.
Universiti Pendidikan Sultan Idris, Malaysia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Guntoro, S.T., M.Kom., MTA.
Universiti Utara Malaysia, Malaysia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Anita Febriani, S.T., M.IT.
Universitas Hang Tuah Pekanbaru, Pekanbaru - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Erni Rouza, S.T., M.Kom.

Universitas Pasir Pengaraian (UPP), Rokan Hulu - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Yumi Novita Dewi, S.Kom., M.Kom.

Universitas Nusa Mandiri, Jakarta - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Moh. Miftakhur Rokhman, S.Si., M.Kom.

Institut Teknologi Nasional (ITN), Malang - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Lisnawita, S.Kom., M.Kom.

Universitas Lancang Kuning, Pekanbaru - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Imam Ahmad, S.Kom., M.Kom.

Universitas Teknokrat Indonesia, Bandar Lampung - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Muhammad Rizal H, S.Kom., M.T.

Universitas Teknologi Akba Makassar, South Sulawesi - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Kelik Sussolaikah, S.Kom., M.Kom.

Universitas PGRI Madiun, East Java - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Dadan Irwan, S.Kom., M.Kom.
Universitas Islam 45, Bekasi, West Java - Indonesia

[Sinta](#) | [Scopus](#) | [Google Scholar](#)



Kementerian Pendidikan Tinggi,
Sains, dan Teknologi

SERTIFIKAT Akreditasi Jurnal



No. SK : 295/C/C3/KPT/2026

Tanggal : 2 Januari 2026

Direktur Penelitian, dan Pengabdian kepada Masyarakat
dengan ini memberikan kepada

***Malcom: Indonesian Journal of Machine
Learning and Computer Science***

EISSN : 27758575
Publisher : Institut Riset dan Publikasi Indonesia

Peringkat Akreditasi Jurnal Ilmiah Periode II 2025

Akreditasi Ulang di Peringkat 3 mulai
Volume 5 Nomor 2 Tahun 2025 sampai Volume 10 Nomor 1 Tahun 2030



I Ketut Adnyana
NIP 196805151994031004

Accreditation (2025-2030)

Vol. 6 No. 2 (2026): MALCOM April 2026

Author Afiliation (Vol. 6 No. 2 April 2026):

DOI: <https://doi.org/10.57152/malcom.v6i2>

Published: 2026-03-26

Articles

Implementasi Internet of Things pada Smart Aquaponics Menggunakan Sensor Level Air dan Sensor BH1750 untuk Monitoring Parameter Lingkungan Berbasis Web Dashboard

Implementation of an Internet of Things-Based Smart Aquaponics Using Water Level and BH1750 Light Sensors with Web-Based Environmental Monitoring

Fera Darmayanti, S. Samsugi

442-452



DOI : <https://doi.org/10.57152/malcom.v6i2.2582>



Abstract View: 644, 

K-Means and Decision Tree Models for Analysis Students Perception of Digital Marketing

Tesdiq Prigel Kaloka, Ahmadi Marta, Mariyam, Nadila Deryza Syofrin

453-461

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2360>  Abstract View: 299, 

Pendekatan Multi-Hirarki untuk Grading Sarang Burung Walet Berdasarkan Deteksi Bentuk dan Klasifikasi Warna

Multi-Hierarchical Machine Learning Approach for Edible Bird's Nest Grading Based on Shape Detection and Color Classification

Reinhard Alfaries Saemani, Danny Manongga, Hendry Hendry

462-474

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2594>  Abstract View: 301, 

Implementation of Machine Learning to Predict The Timeliness of Graduation of Employees on Study Assignment at Company X

Parenda Rizkya Permata, Imam Yuadi

475-482

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2474>  Abstract View: 199, 

Comparative Analysis of ANN, 1D-CNN, and LSTM for Multi-Label Action Prediction in IoT-Based Hydroponic Control Systems

Reza Octaviany, Suhendro Yusuf Irianto

483-493

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2581>  Abstract View: 197, 

Predicting Students' Mathematics Scores from Reading Scores Using Supervised Learning

Nofita Fitriyani

494-503



 DOI : <https://doi.org/10.57152/malcom.v6i2.2555>  Abstract View: 211, 

Comparative Analysis of Random Forest, Explainable Boosting Machine and Ensemble Stacking Performance for Hepatitis C Disease Classification

Anastasia Ngeni Bagur, Irfan Pratama

504-512



 DOI : <https://doi.org/10.57152/malcom.v6i2.2561>  Abstract View: 172, 

Implementasi Perbandingan YOLO v8, v9, dan v11 dalam Penerapan Tata Tertib K3: Deteksi Penggunaan Helm Keselamatan di Lingkungan Konstruksi

Implementation and Comparison of YOLO v8, v9, and v11 in Occupational Safety Regulations:
Detecting Safety Helmet Usage in Construction Environments

Rafael Praseli, Nenden Siti Fatonah, Diah Aryani, Hani Dewi Ariessanti

513-523



 DOI : <https://doi.org/10.57152/malcom.v6i2.2554>  Abstract View: 474, 

UI/UX Desain Mobile Apps Skin Check untuk Identifikasi Jenis Kulit dan Rekomendasi Produk Skincare

UI/UX Design of Mobile Apps Skin Check to Identify Skin Type and Recommend Skincare Products

Nava Gia Ginasta, Fariz Faqih Firdaus

524-533



 DOI : <https://doi.org/10.57152/malcom.v6i2.2383>  Abstract View: 230, 

Analisis Sentimen Berbasis Aspek pada Ulasan Pariwisata Menggunakan Varian Algoritma K-Nearest Neighbor

Aspect-Based Sentiment Analysis of Tourism Reviews Using Variants of the K-Nearest Neighbor Algorithm

Anastasya Nurfitriyani Hidayat, Ahmad Luky Ramdani, Luluk Muthoharoh

534-545

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2522>  Abstract View: 385, 

Unsupervised Text Mining of Employee Feedback for Identifying Organizational Strengths and Improvement Areas

Febri Ari Wicaksono, Imam Yuadi

546-554

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2482>  Abstract View: 148, 

Systematic Review and Bibliometric Mapping on Image Processing in Electronic Health Records

Amir Hamzah Dinnillah, Fikri Maulana

555-567

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2462>  Abstract View: 146, 

Time Series Approach for Analysis and Prediction of Malware Trends Based on Open Source Intelligence

Tommy Nugraha Manoppo, Abdul Gani Fadhlulrahman, Yudi Prayudi

568-576



DOI : <https://doi.org/10.57152/malcom.v6i2.2580> Abstract View: 163,

Systematic Literature Review of Transfer Learning for Pneumonia Classification in Chest X-Rays

Erlan Bachtiar, Amir Hamzah Dinnillah, Yan Rianto

577-590



DOI : <https://doi.org/10.57152/malcom.v6i2.2470> Abstract View: 1520,

The Lexicon-Based Sentiment Analysis for Stock Price Prediction in Islamic Fashion Industry

Entis Sutisna; Rahma Dafitri; Dian Daryani

591-598



DOI : <https://doi.org/10.57152/malcom.v6i2.2586> Abstract View: 148,

Performance Comparison of Facial Skin Type Classification Using the Segment Anything Model

Nisrina Nur Kumala, Asfan Muqtadir, Amaludin Arifia

599-607



DOI : <https://doi.org/10.57152/malcom.v6i2.2604> Abstract View: 402,

Implementasi Algoritma K-Nearest Neighbors dalam Analisis Sentimen Ulasan Aplikasi Bank Aladin

Implementation of the K-Nearest Neighbors Algorithm for Sentiment Analysis on Aladin Bank Application Reviews

Taniya Raisha Dwiva Putri, M. Rudi Sanjaya, MGS Afriyan Firdaus, Dwi Rosa Indah

608-618



 DOI : <https://doi.org/10.57152/malcom.v6i2.2633>  Abstract View: 334, 

Pengaruh ChatGPT Terhadap Peforma Akademik Mahasiswa Menggunakan Algoritma Support Vector Machine

The Impact of ChatGPT on Students' Academic Performance Using the Support Vector Machine Algorithm

Gita Aulia, Sriani Sriani

619-631

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2630>  Abstract View: 274, 

Analisis Sentimen Ulasan Pengguna Aplikasi Akupunktur di Play Store Menggunakan Naïve Bayes

Sentiment Analysis User Reviews of Acupuncture Applications on Play Store Using Naïve Bayes

Wahyu Teja Kusuma, Leny Candra Kurniawan, Nindynar Rikatsih, Puspo Wardoyo

632-639

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2620>  Abstract View: 184, 

Analisis Perbandingan Kinerja IndoBERT dan TF-IDF dalam Mengklasifikasikan Sentimen EDOM Menggunakan Algoritma K-Nearest Neighbor

Comparative Analysis of IndoBERT and TF-IDF Performance in Classifying EDOM Sentiments Using the K-Nearest Neighbor Algorithm

Tifanny Nabarian, Siti Nurhalizah, Salman El Farisi

640-650

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2614>  Abstract View: 642, 

Desain dan Implementasi Sistem Presensi Portabel Berbasis Internet of Things dengan Validasi Multi-Faktor untuk Akuntabilitas Tenaga Kerja Lapangan

Design and Implementation of a Portable Internet of Things-Based Attendance System with Multi-Factor Validation for Field Workforce Accountability

Marhaeni Marhaeni, Riadi Marta Dinata, Agus Sofwan, Dikky Suryadi, Niko Purnomo

651-662

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2618>  Abstract View: 169, 

Implementasi Metode Hybrid Backpropagation Neural Network dan Particle Swarm Optimization untuk Prediksi Konsumsi Listrik Rumah Tangga Berdasarkan Golongan Tarif

Implementation of Hybrid Backpropagation Neural Network and Particle Swarm Optimization for Predicting Household Electricity Consumption Based On Tariff Categories

Yuni Artha Chyntia Saragih, Erma Suryani

663-673

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2519>  Abstract View: 164, 

Segmentasi Pelanggan Kartu Halo Telkomsel Berbasis K-Means di Wilayah Sumbagsel

K-Means-Based Customer Segmentation of Telkomsel Kartu Halo in the Sumbagsel

Muhammad Bayu Samudra, Alifa Putri Shahabiyah, Aliya Faiza, Edo Wicaksono, Ken Ditha Tania, Allsella Meiriza, Ahmad Rifai

674-683

 PDF

 DOI : <https://doi.org/10.57152/malcom.v6i2.2607>  Abstract View: 236, 

Penentuan Metode MOORA pada Sistem Pendukung Keputusan Penentuan Tempat Praktik Kerja Lapangan

Determination of the MOORA Method in the Decision Support System for Determining Field Work Practice Location

Zilan Nazmi Dalimunthe, Arridha Zikra Syah, Febby Madonna Yuma

684-694



 DOI : <https://doi.org/10.57152/malcom.v6i2.2599>  Abstract View: 219, 

Penerapan Model Recency, Frequency, Monetary untuk Segmentasi Pola Perilaku Pelanggan Indibiz

Application of the Recency, Frequency, Monetary Model for Segmentation of Indibiz Customer Behavior Patterns

Aulia Pinkasari, Meiyin Monica Amilia Putri, Gibril Abdurahman, Ahmad Fadhil Rizqi, Ken Ditha Tania, Allsela Meiriza, Ahmad Rifai

695-704



 DOI : <https://doi.org/10.57152/malcom.v6i2.2611>  Abstract View: 214, 

Perbandingan Model Klasifikasi Supervised Machine Learning dalam Knowledge Discovery Layanan TI Pertamina Prabumulih

Comparison of Supervised Machine Learning Classification Models in Knowledge Discovery of Pertamina Prabumulih IT Services

Shafa Aurelliza Arian, Putri Rahel Alifia, Bagus Prihantoro, Muhammad Iqbal Disriansyah, Ken Ditha Tania, Alsella Meiriza, Ahmad Rifai

705-715



 DOI : <https://doi.org/10.57152/malcom.v6i2.2617>  Abstract View: 256, 

Penerapan Database Tuning Melalui Indexing dan Optimasi Query pada Sistem Kuesioner

Akademik

Implementation of Database Tuning Through Indexing and Query Optimization in an Academic Questionnaire System

Ikhwan Ikhwan, Angga Pramadjaya

716-725



DOI : <https://doi.org/10.57152/malcom.v6i2.2610>



Abstract View: 440,

Tinjauan Literatur Sistematis Tentang Algoritma Deteksi Objek untuk Sistem Otonom: Evaluasi Arsitektur CNN, Tantangan Dataset, dan Strategi Implementasi

Systematic Literature Review on Object Detection Algorithms for Autonomous Systems: Evaluation of CNN Architectures, Dataset Challenges, and Implementation Strategies

Mohammad Yasser Arafat, Muhammad Bagus Andra

726-736



DOI : <https://doi.org/10.57152/malcom.v6i2.2584>



Abstract View: 203,

Performance Evaluation of the DHT11 Sensor for MATLAB-Based Temperature and Humidity Measurements

Muhammad Chusni Marzuki, Ulul Ilmi, Arif Budi Laksono, Eko Wahyu Santoso, Abdur Rohman Wakhid

737-749



DOI : <https://doi.org/10.57152/malcom.v6i2.2565>



Abstract View: 331,

Segmentasi Spasio-Temporal Performa Penjualan Produk Sepatu Menggunakan Algoritma K-Means

Spatio-temporal Segmentation of Footwear Sales Performance Using the K-Means Algorithm

Shindy Yuliyatini, Eva Yulyanti, Gandung Triyono

750-761



 DOI : <https://doi.org/10.57152/malcom.v6i2.2609>  Abstract View: 241, 

Estimasi Tax Gap Indonesia: Analisis Diskrepansi PDRB Resmi Berdasarkan Nighttime Lights Berbasis Machine Learning

Estimating Indonesia's Tax Gap: Analysis of Discrepancies in Official GRDP Based on Machine Learning-Based Nighttime Lights

Fachri Husein Harahap

762-774



 DOI : <https://doi.org/10.57152/malcom.v6i2.2568>  Abstract View: 519, 

Adoption Without Optimization: Text Mining and Business Intelligence Analysis on Culinary SME Digital Literacy

Rahma Dafitri, Dian Daryani, Entis Sutisna

775-787



 DOI : <https://doi.org/10.57152/malcom.v6i2.2622>  Abstract View: 121, 

Sistem Analitik Umpan Balik YouTube Berbasis Big Data dan Generative AI

Big Data-Based YouTube Feedback Analytics Architecture with Generative AI Integration

Asro Asro, Novi Rukhviyanti

788-799



 DOI : <https://doi.org/10.57152/malcom.v6i2.2626>  Abstract View: 212, 

Analisis Sentimen Analisis Sentimen Ulasan Aplikasi Telemedicine di Indonesia Menggunakan XLM-RoBERTa dan Non-Negative Matrix Factorization untuk Mendukung Sustainable Development Goal 3

Sentiment Analysis of Telemedicine App Reviews in Indonesia Using XLM-RoBERTa and Non-Negative Matrix Factorization for Supporting Sustainable Development Goal 3

Asha Sembiring, M. Imam Santoso, Alfian Ramadhan Sembiring

800-809



DOI : <https://doi.org/10.57152/malcom.v6i2.2621> Abstract View: 323,

Multivariat Support Vector Regression untuk Peramalan Permintaan Suku Cadang Intermittent pada Industri Perkeretaapian

Multivariate Support Vector Regression for Intermittent Spare Parts Demand Forecasting in the Railway Industry

Syayid Al Afghoni, Iwan Vanany

810-820



DOI : <https://doi.org/10.57152/malcom.v6i2.2549> Abstract View: 178,

Penerapan Data Mining untuk Prediksi Permintaan Obat Menggunakan Algoritma Regresi Linier

Implementation of Data Mining for Drug Demand Prediction Using Linear Regression Algorithm

Hafiz Pramuji Almi, Hambali Hambali, Ari Dermawan

821-828



DOI : <https://doi.org/10.57152/malcom.v6i2.2615> Abstract View: 154,

Analisis Ulasan Sentimen pada Aplikasi Mobile BSB Menggunakan Algoritma Convolutional

Neural Networks dan Support Vector Machine

Sentiment Review Analysis on the BSB Mobile Application Using Convolutional Neural Networks and Support Vector Machine Algorithm

Muhamamad Faza, M. Rudi Sanjaya, Ali Ibrahim, M. Husni Syahbani

829-838



DOI : <https://doi.org/10.57152/malcom.v6i2.2637>  Abstract View: 176, 

Analisis Komparatif Arsitektur MobileNet untuk Klasifikasi Tipe Jerawat pada Citra Wajah

Comparative Analysis of MobileNet Architectures for Acne Type Classification on Facial Images

Sukma Syafika, Susi Marianingsih

839-851



DOI : <https://doi.org/10.57152/malcom.v6i2.2643>  Abstract View: 369, 

Spatial Modeling of Highly Dispersed Poverty Data in East Java Using the Geographically Weighted Negative Binomial Regression Method

Nova Putri Dwi Sugiarti, Trimono Trimono, Shindi Shella May Wara

852-864



DOI : <https://doi.org/10.57152/malcom.v6i2.2636>  Abstract View: 108, 

Implementation of Time Series Forecasting for Inflation Prediction in Indonesia

Mustakim Mustakim, Wahyu Eka Putra, Hartono Hartono

865-874



DOI : <https://doi.org/10.57152/malcom.v6i2.2172>  Abstract View: 339, 

Klasifikasi Jenis Sayuran Berbasis Citra Menggunakan Convolutional Neural Network dengan Penyediaan Informasi Kandungan Nutrisi

Image-Based Vegetable Type Classification Using Convolutional Neural Network with Nutritional Content Information Provision

Intan Idaman Halawa, Choirul Basir

875-885



DOI : <https://doi.org/10.57152/malcom.v6i2.2267>



Abstract View: 388,

Comparative Sentiment Analysis of TheoTown Reviews on Steam and Google Play Store Using Support Vector Machine

Shanda Aura, Dian Novianto

886-896



DOI : <https://doi.org/10.57152/malcom.v6i2.2661>



Abstract View: 285,

Implementasi Model Klasifikasi Decision Tree untuk Evaluasi Kinerja Organisasi dan Dukungan Pengambilan Keputusan

Implementation of Decision Tree Classification Model for Organizational Performance Evaluation and Decision Support

Tamara Juliyanti, Xenia Clarissa Valencia Marpaung, Nanda Salsabila Az-Zahra, Silvia Nabilaputri, Allsela Meiriza, Ken Ditha Tania, Ahmad Rifai

897-907



DOI : <https://doi.org/10.57152/malcom.v6i2.2646>



Abstract View: 214,

Perancangan Tata Kelola TI Laboratorium Internet of Things PNUP Menggunakan Kerangka COBIT 2019

IT Governance Design for PNUP Internet of Things Laboratory Using the COBIT 2019 Framework

Amhar Davi Dewantara, Iin Karmila Yusri, Budy Santoso

908-922



DOI : <https://doi.org/10.57152/malcom.v6i2.2668> Abstract View: 178,

Perbandingan Support Vector Machine dan Naïve Bayes untuk Klasifikasi Sentimen Ulasan E-Commerce

Comparison of Support Vector Machine and Naïve Bayes for E-Commerce Review Sentiment Classification

Dita Nurmadewi, Zakiul Fahmi Jailani, Haris Rafi, Dimas Aryo Anggoro, Dewi Setiowati

923-933



DOI : <https://doi.org/10.57152/malcom.v6i2.2648> Abstract View: 679,

Evaluasi Komparatif Kinerja Algoritma Naïve Bayes dan Logistic Regression dalam Klasifikasi Sentimen Komentar YouTube Berbahasa Indonesia

Comparative Performance Evaluation of Naïve Bayes and Logistic Regression for Indonesian YouTube Comment Sentiment Classification

Meisa Monica, Asto Purwanto

934-943



DOI : <https://doi.org/10.57152/malcom.v6i2.2639> Abstract View: 461,

Konsep Lean Supply Chain dalam Meningkatkan Efisiensi Operasional

The Lean Supply Chain Concept in Improving Operational Efficiency

Rian Refanza Lumbantoruan, Jeperson Hutahaeen, Parini Parini

944-951



DOI : <https://doi.org/10.57152/malcom.v6i2.2606> Abstract View: 162,

Development of a Portable Smart Feeder Based on Scheduling and Precision Dosing for Fish Feeding in Aquaculture

Arif Dwi Kuncoro, Tri Widodo, Yuri Rahmanto, Sahrial Ihsani Ishak

952-961



DOI : <https://doi.org/10.57152/malcom.v6i2.2559> Abstract View: 154,

Development and Optimization of YOLOv12 for Autonomous Vehicle Navigation Systems

Dian Ramadhani, Muhammad Muttakin, Hidayat Hatta Irsyad, Edi Susilo, Rahmat Rizal Andi

962-972



DOI : <https://doi.org/10.57152/malcom.v6i2.2634> Abstract View: 205,

Intelligent Automation of Linux System Administration via the NeuroSysAI Autonomous Agent: Security and Cost-Efficiency Analysis

Auliya Ur Rahman Ash-Shidqi, Helmy Faisal Muttaqin

973-985



DOI : <https://doi.org/10.57152/malcom.v6i2.2588> Abstract View: 598,

Penerapan Squeeze-and-Excitation Attention pada DenseNet169 untuk Klasifikasi Multi-Kelas Citra X-Ray Dada

Squeeze-and-Excitation Attention on DenseNet169 for Multi-Class Classification of Chest X-Ray Images

Lyra Zulyanda Daulay, Benny Sukma Negara, Yelfi Vitriani, Iwan Iskandar, Fitra Kurnia

986-996



Spatio-Temporal Segmentation of Footwear Sales Performance Using the K-Means Algorithm

Segmentasi Spasio-Temporal Performa Penjualan Produk Sepatu Menggunakan Algoritma K-Means

Eva Yulyanti¹, Shindy Yuliyatini^{2*}, Gandung Triyono³

^{1,2,3}Magister Ilmu Komputer, Universitas Budi Luhur, Jakarta, Indonesia

E-Mail: ¹evayulyanti02@gmail.com,

²shindy7yuliyatini@gmail.com ³gandung.triyono@budiluhur.ac.id

Received Feb 17th 2026; Revised Mar 25th 2026; Accepted Mar 30th 2026; Available Online Apr 24th 2026

Corresponding Author: Shindy Yuliyatini

Copyright ©2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Regional sales performance disparities are commonly analyzed using annual aggregation, which may obscure market dynamics. This study proposes a spatio-temporal segmentation model for footwear sales performance using the K-Means algorithm with a State $\times$ Quarter unit of analysis. The Dataset was obtained from the Kaggle platform (Adidas US Sales Dataset) for the 2020–2021 period and aggregated from 400 transactions into 181 observations using total sales, units sold, operating profit, and operating margin variables. The method includes preprocessing using IQR and winsorizing, Min-Max normalization, and K-Means clustering. Evaluation was conducted using the Silhouette Score, Davies–Bouldin Index (DBI), Calinski–Harabasz Index (CHI), and stability testing with the Adjusted Rand Index (ARI) as an additional validation method to assess clustering consistency. The results indicate three clusters with distributions of 55.8% (low), 32.0% (medium), and 12.2% (high). A Silhouette Score of 0.4263 indicates moderate cluster separation, while an ARI of 1.000 demonstrates very high stability. Spatio-temporal analysis shows that 52% of states experienced changes in clustering across quarters, indicating regional performance dynamics not captured by annual aggregation. The scientific contribution of this study lies in applying spatio-temporal segmentation using K-Means, with comprehensive cluster quality and stability evaluation, to more accurately identify the dynamics of regional distribution performance.

Keywords: Cluster Stability, Footwear Sales, K-Means Clustering, Silhouette core, Spatio-Temporal Segmentation

Abstrak

Ketimpangan performa penjualan antarwilayah umumnya dianalisis melalui agregasi tahunan yang berpotensi menyembunyikan dinamika pasar. Penelitian ini bertujuan mengusulkan model segmentasi performa penjualan sepatu berbasis *spatio-temporal* menggunakan algoritma K-Means dengan unit analisis *State $\times$ Quarter*. Dataset diperoleh dari platform Kaggle (Adidas US Sales Dataset) periode 2020-2021 dan diagregasi dari 400 transaksi menjadi 181 observasi menggunakan variabel *total sales*, *units sold*, *operating profit*, dan *operating margin*. Metode meliputi *preprocessing* menggunakan IQR dan *winsorizing*, normalisasi *Min-Max*, serta *clustering K-Means*. Evaluasi dilakukan menggunakan *Silhouette Score*, *Davies-Bouldin Index (DBI)*, *Calinski-Harabasz Index (CHI)*, dan uji stabilitas *Adjusted Rand Index (ARI)* sebagai metode validasi tambahan untuk mengukur konsistensi hasil *clustering*. Hasil menunjukkan tiga *cluster* dengan distribusi 55,8% (rendah), 32,0% (menengah), dan 12,2% (tinggi). Nilai *Silhouette Score* sebesar 0,4263 menunjukkan pemisahan *cluster* moderat, sedangkan ARI sebesar 1,000 menunjukkan stabilitas sangat tinggi. Analisis spasio-temporal menunjukkan 52% *state* mengalami perubahan *cluster* antarkuartal. Temuan ini menunjukkan adanya dinamika performa wilayah yang tidak terdeteksi dalam agregasi tahunan. Kontribusi ilmiah penelitian ini adalah penerapan segmentasi *spatio-temporal* berbasis K-Means dengan evaluasi kualitas dan stabilitas *cluster* untuk mengidentifikasi dinamika performa distribusi wilayah secara lebih akurat.

Kata Kunci: K-Means Clustering, Penjualan Sepatu, Segmentasi Spatio-Temporal, Silhouette Score, Stabilitas Cluster

1. PENDAHULUAN

Periode 2020-2021 merupakan fase yang ditandai oleh dinamika signifikan dalam industri ritel sepatu (*footwear retail*) di Amerika Serikat, termasuk fluktuasi permintaan, gangguan rantai pasok, serta perubahan

pola konsumsi antarwilayah. Perusahaan dengan jaringan distribusi lintas 50 *state* menghadapi tantangan dalam menjaga keseimbangan distribusi dan profitabilitas karena performa penjualan antarwilayah dapat berubah dalam interval waktu yang relatif singkat. Variasi ini menunjukkan bahwa performa distribusi tidak bersifat homogen, melainkan dipengaruhi oleh faktor spasial dan temporal yang saling berinteraksi. Untuk menjawab kompleksitas tersebut, penelitian ini mengusulkan pendekatan segmentasi berbasis unit analisis *State × Quarter* yang secara simultan mengintegrasikan dimensi spasial dan temporal dalam analisis performa distribusi. Pendekatan ini memungkinkan identifikasi perubahan performa wilayah secara periodik sehingga dinamika jangka pendek yang sering tereduksi dalam agregasi tahunan dapat diungkap secara lebih akurat. *Dataset* penelitian terdiri dari 400 transaksi yang ditransformasikan menjadi 181 observasi berbasis *State × Quarter* dengan variabel *total sales*, *units sold*, *total profit*, dan *operating margin*. Data multivariat ini menuntut metode analisis yang mampu mengungkap pola tersembunyi secara objektif, sehingga pendekatan data *mining* menjadi relevan untuk digunakan. Hal ini didukung oleh penelitian [1][2] yang menunjukkan efektivitas teknik data *mining* dalam analisis penjualan.

Salah satu metode yang umum digunakan dalam analisis data adalah algoritma *K-Means*, yang mengelompokkan data berdasarkan kedekatan karakteristik numerik secara efisien. Dalam konteks penjualan, metode ini digunakan untuk mengidentifikasi pola kebutuhan produk sebagaimana ditunjukkan dalam [3]. Selain itu, *K-Means* banyak digunakan dalam segmentasi pelanggan, seperti pada [4] [5]. Penerapan serupa juga ditemukan dalam [6], yang menunjukkan bahwa *Clustering* mampu mengidentifikasi produk unggulan berdasarkan pola penjualan. Dalam konteks pemasaran, Algoritma *Clustering* berperan penting dalam mendukung strategi bisnis berbasis data. Penelitian [7] menunjukkan bahwa segmentasi pelanggan dapat meningkatkan efektivitas strategi pemasaran. Temuan ini diperkuat oleh [8], yang mengintegrasikan *Clustering* dengan pendekatan kecerdasan buatan. Selain itu, studi [9][10] menunjukkan bahwa *Clustering* mampu mengidentifikasi pola perilaku konsumen dalam berbagai konteks pasar. Pendekatan berbasis perilaku seperti dalam [11] juga memperkuat relevansi metode ini dalam analisis transaksi.

Meskipun demikian, sebagian besar penelitian tersebut masih berfokus pada entitas pelanggan atau produk. Beberapa penelitian mulai mengarah pada analisis distribusi, seperti [12], namun belum mengintegrasikan dimensi waktu secara eksplisit. Penelitian lain seperti [3] juga masih terbatas pada pendekatan statis. Dalam konteks industri sepatu, penelitian [13][14] menunjukkan penerapan *Clustering* dalam segmentasi pasar, tetapi belum mempertimbangkan dinamika wilayah berbasis waktu. Di sisi lain, perkembangan literatur menunjukkan bahwa integrasi dimensi spasial dan temporal menjadi semakin penting dalam analisis data kompleks. Hal ini terlihat dalam penelitian [15] yang memanfaatkan data spasio-temporal untuk analisis distribusi. Pendekatan serupa juga dikembangkan dalam [16] yang menekankan pentingnya dimensi waktu dalam analisis wilayah. Selain itu [17] menunjukkan bahwa algoritma *Clustering* dapat diperluas untuk mengakomodasi struktur data spasio-temporal. Pendekatan yang lebih kompleks juga diperkenalkan dalam [16], yang menegaskan bahwa pola data bersifat dinamis. Penelitian lain seperti [18] juga memperkuat pentingnya analisis berbasis spasial dalam memahami distribusi fenomena.

Selain aspek metode, kualitas hasil *Clustering* sangat dipengaruhi oleh proses evaluasi. Penelitian [19] menunjukkan bahwa pemilihan jumlah *cluster* yang optimal merupakan faktor penting dalam menghasilkan segmentasi yang valid dan dapat diinterpretasikan dengan baik. Oleh karena itu, masih terdapat kesenjangan penelitian yang signifikan dalam literatur yang ada. Sebagian besar penelitian sebelumnya cenderung berfokus pada segmentasi pelanggan atau produk, sehingga belum mampu merepresentasikan performa distribusi berbasis wilayah secara komprehensif. Di sisi lain, penelitian berbasis penjualan umumnya menggunakan pendekatan agregatif tahunan yang berpotensi mengabaikan dinamika temporal jangka pendek. Selain itu, meskipun pendekatan spasio-temporal telah berkembang dalam berbagai bidang, penerapannya dalam konteks distribusi ritel masih terbatas dan belum banyak dikombinasikan dengan teknik *Clustering* berbasis *K-Means*. Dengan demikian, belum terdapat penelitian yang secara simultan mengintegrasikan algoritma *K-Means* dengan unit analisis spasial (*state*) dan temporal (*quarter*) untuk menganalisis performa distribusi ritel dalam skala nasional.

Berdasarkan telaah literatur tersebut, terdapat tiga kesenjangan utama, yaitu dominasi penelitian *Clustering* pada segmentasi pelanggan atau produk, penggunaan pendekatan agregatif statis yang belum menangkap dinamika temporal jangka pendek, serta keterbatasan integrasi algoritma *K-Means* dengan unit analisis spasio-temporal dalam konteks distribusi ritel berskala nasional. Berdasarkan kesenjangan tersebut, penelitian ini menghasilkan model segmentasi performa wilayah berbasis *State × Quarter* menggunakan algoritma *K-Means* dengan evaluasi *cluster* yang komprehensif. Pendekatan ini memungkinkan identifikasi pola performa wilayah secara dinamis serta perubahan posisi relatif antar wilayah dalam interval kuartalan. Penelitian ini bertujuan mengembangkan segmentasi berbasis integrasi spasial-temporal dan menentukan konfigurasi *cluster* optimal pada 50 *state* di Amerika Serikat. Kontribusi penelitian terletak pada penerapan pendekatan *State × Quarter* yang mengintegrasikan dimensi spasial dan temporal secara simultan. Secara teoretis, penelitian ini memperluas penerapan *K-Means* pada segmentasi performa distribusi wilayah berskala nasional, sedangkan secara praktis memberikan dasar analitis untuk pengambilan keputusan distribusi yang lebih responsif dan berbasis data.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis segmentasi data untuk mengelompokkan performa distribusi wilayah dengan algoritma *K-Means*. Unit analisis ditetapkan dalam bentuk agregasi *State* × *Quarter* guna mengintegrasikan dimensi spasial dan temporal. Tahapan penelitian meliputi: pengambilan data transaksi, transformasi dan agregasi data, seleksi variabel, *preprocessing* (*cleaning*, deteksi *outlier*, *winsorizing*, dan normalisasi), proses *Clustering*, penentuan jumlah *cluster* optimal menggunakan validasi internal, uji stabilitas hasil, serta analisis dinamika keanggotaan *cluster* antar kuartal. Rangkaian tahapan tersebut ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Pengambilan dan Agregasi Data

Data penelitian diperoleh dari platform Kaggle berupa transaksi penjualan sepatu Adidas di Amerika Serikat periode 2020-2021 menggunakan Adidas US Sales Dataset yang tersedia secara publik. Dataset berbentuk *transaction-level* dengan 13 atribut: *Retailer*, *Retailer Id*, *Invoice Date*, *Region*, *State*, *City*, *Product*, *Price per Unit*, *Units Sold*, *Total Sales*, *Operating Profit*, *Operating Margin*, dan *Sales Method*. Untuk analisis *spasio-temporal*, data ditransformasi ke unit analisis *State* × *Quarter* sehingga setiap observasi merepresentasikan performa distribusi suatu *state* per kuartal. Proses agregasi dilakukan dengan menjumlahkan *Total Sales*, *Units Sold*, dan *Operating Profit* sebagai indikator volume kumulatif, serta menghitung rata-rata *Operating Margin* sebagai representasi efisiensi relatif pada setiap kombinasi *State* × *Quarter*. Secara teoritis terbentuk 400 observasi (50 *state* × 8 kuartal), namun setelah *preprocessing* jumlah akhir menjadi 181 observasi valid karena tidak semua *state* memiliki transaksi pada setiap kuartal.

2.2. Seleksi Variabel

Dari 13 atribut transaksi, penelitian ini memilih empat variabel utama untuk merepresentasikan performa distribusi wilayah, yaitu *Total Sales*, *Units Sold*, *Operating Profit*, dan *Operating Margin*. *Total Sales* mencerminkan skala penjualan, *Units Sold* menggambarkan volume permintaan, *Operating Profit* menunjukkan keuntungan absolut, sedangkan *Operating Margin* merepresentasikan efisiensi operasional. Keempat variabel bersifat kuantitatif kontinu sehingga sesuai dengan algoritma *K-Means* berbasis jarak *Euclidean*, serta dipilih dengan prinsip non-redundansi untuk menghindari dominasi dimensi tertentu. Kombinasi variabel ini memungkinkan pembentukan *cluster* yang menangkap heterogenitas performa wilayah secara komprehensif dari sisi skala penjualan dan efisiensi operasional.

2.3. Data Cleaning

Tahap *preprocessing* diawali dengan data *cleaning* untuk memastikan kualitas dataset sebelum *Clustering*. Pemeriksaan dilakukan terhadap *missing value*, duplikasi kombinasi *State* × *Quarter*, serta inkonsistensi struktur data. Hasil pemeriksaan menunjukkan tidak terdapat nilai hilang maupun duplikasi sehingga dataset dinyatakan valid. Selanjutnya, dilakukan deteksi *outlier* karena variabel penjualan dan profit berpotensi memiliki distribusi tidak simetris akibat perbedaan kapasitas pasar antarstate.

2.4. Deteksi Outlier (IQR)

Deteksi *outlier* dilakukan menggunakan metode *Interquartile Range* (IQR). Nilai IQR dihitung sebagai selisih kuartil ketiga dan kuartil pertama sebagaimana Persamaan (1).

$$IQR = Q_3 - Q_1 \quad (1)$$

Berdasarkan nilai IQR tersebut, batas bawah dan batas atas *outlier* ditentukan menggunakan Persamaan (2) dan Persamaan (3).

$$Lower\ Bound = Q_1 - 1.5(IQR) \quad (2)$$

$$Upper\ Bound = Q_3 + 1.5(IQR) \quad (3)$$

Observasi yang berada di luar rentang tersebut dikategorikan sebagai *outlier* dan dilanjutkan ke tahap *winsorizing*.

2.5. Winsorizing

Nilai *outlier* yang teridentifikasi kemudian disesuaikan menggunakan teknik *winsorizing* agar distribusi data tetap stabil tanpa mengurangi ukuran sampel. Secara matematis proses ini ditunjukkan pada Persamaan (4).

$$x_i^* = \begin{cases} \text{Lower Bound jika } x_i < \text{Lower Bound} \\ \text{Upper Bound jika } x_i > \text{Upper Bound} \\ x_i \text{ jika } \text{Lower Bound} \leq x_i \leq \text{Upper Bound} \end{cases} \quad (4)$$

Pendekatan ini mengurangi sensitivitas algoritma *K-Means* terhadap nilai ekstrem sekaligus mempertahankan variasi performa antarwilayah.

2.6. Normalisasi

Setelah penyesuaian *outlier*, dilakukan normalisasi untuk menyamakan skala antarvariabel. Hal ini diperlukan karena *K-Means* berbasis jarak Euclidean sensitif terhadap perbedaan magnitudo. Normalisasi menggunakan *Min-Max Scaling* ke rentang [0,1] sebagaimana Persamaan (5):

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

Tahap ini menghasilkan data dengan skala sebanding sehingga setiap variabel berkontribusi proporsional dalam pembentukan *cluster*.

2.7. Penentuan rentang k (2-10)

Penentuan jumlah *cluster* (*k*) merupakan parameter penting dalam *K-Means* karena memengaruhi tingkat granularitas segmentasi. Nilai *k* yang terlalu kecil menyebabkan *under-segmentation*, sedangkan nilai yang terlalu besar berisiko menurunkan stabilitas dan interpretabilitas. Dalam penelitian ini, rentang *k* ditetapkan antara 2 hingga 10. Batas bawah *k* = 2 digunakan sebagai konfigurasi minimal untuk membedakan karakteristik performa distribusi, sedangkan batas atas *k* = 10 dipilih untuk menghindari fragmentasi berlebihan mengingat jumlah observasi terbatas dan tujuan segmentasi bersifat manajerial. Seluruh nilai *k* dalam rentang tersebut diuji menggunakan *K-Means* dengan inisialisasi *K-Means++* dan *n_init* = 50, kemudian dievaluasi menggunakan indeks validasi internal.

2.8. K-Means Clustering

Setelah jumlah *cluster* optimal diperoleh, algoritma *K-Means* digunakan untuk membentuk segmentasi akhir. Variabel yang digunakan berupa indikator kuantitatif performa distribusi (*Total Sales*, *Units Sold*, *Operating Profit*, dan *Operating Margin*). Secara matematis, *K-Means* meminimalkan variasi *intra-cluster* melalui *Within-Cluster Sum of Squares* (WCSS) sebagaimana Persamaan (6).

$$WCSS = \sum_{k=1}^K \sum_{x_i \in C_k} |x_i - \mu_k|^2 \quad (6)$$

Di mana x_i adalah observasi ke-*i*, C_k *cluster* ke-*k*, dan μ_k *centroid cluster*. Jarak yang digunakan adalah *Euclidean distance*. Inisialisasi *centroid* menggunakan *K-Means++* dengan *n_init* = 50 untuk meningkatkan stabilitas dan mengurangi kemungkinan solusi lokal. Parameter ini bertujuan menghasilkan konfigurasi *cluster* yang konsisten dan representatif terhadap struktur data.

2.9. Evaluasi Internal

Evaluasi internal digunakan untuk menentukan jumlah *cluster* optimal berdasarkan kualitas pemisahan dan kekompakan *cluster* tanpa memerlukan label eksternal, sehingga hanya bergantung pada struktur intrinsik data. Prosesnya meliputi: (1) menetapkan rentang *k* = 2–10, (2) menjalankan *K-Means* untuk setiap *k*, (3) menghitung indeks evaluasi, (4) membandingkan hasil, dan (5) memilih *k* terbaik berdasarkan kualitas statistik dan interpretabilitas. Tiga indeks digunakan, yaitu *Silhouette Score*, *Calinski-Harabasz Index* (CHI), dan *Davies-Bouldin Index* (DBI), karena memberikan perspektif evaluasi yang saling melengkapi. *Silhouette Score* mengukur kesamaan objek terhadap *cluster*-nya dibandingkan *cluster* lain dengan nilai -1 hingga 1, di mana nilai lebih tinggi menunjukkan kualitas *cluster* lebih baik. Secara matematis dihitung dengan Persamaan (7).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (7)$$

Di mana *a(i)* adalah rata-rata jarak objek terhadap anggota dalam *cluster* yang sama dan *b(i)* adalah jarak rata-rata minimum terhadap *cluster* lain terdekat [19]. *Calinski-Harabasz Index* (CHI) mengevaluasi

rasio variasi antar-*cluster* terhadap variasi *intra-cluster*. Nilai yang lebih tinggi menunjukkan pemisahan *cluster* yang lebih baik. CHI dihitung menggunakan Persamaan (8).

$$CH = \frac{\text{Between-Cluster Dispersion}/(k-1)}{\text{Within-Cluster Dispersion}/(n-k)} \quad (8)$$

Di mana B_k merepresentasikan dispersi antar-*cluster*, W_k dispersi *intra-cluster*, k jumlah *cluster*, dan n jumlah observasi. Davies–Bouldin Index (DBI) mengukur kesamaan antar *cluster* berdasarkan rasio dispersi *intra-cluster* dan jarak antar *centroid*. Semakin kecil nilai DBI, semakin baik kualitas *cluster*. Formulasinya ditunjukkan pada Persamaan (9).

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right) \quad (9)$$

Di mana S_i dan S_j adalah rata-rata jarak *intra-cluster*, dan M_{ij} adalah jarak antar-*centroid cluster*. Jumlah *cluster* optimal ditentukan dengan memilih nilai k yang menghasilkan Silhouette dan CHI tinggi serta DBI rendah secara simultan, dengan mempertimbangkan interpretabilitas manajerial.

2.10. Uji Stabilitas

Uji stabilitas dilakukan untuk memastikan hasil segmentasi tidak bergantung pada variasi inisialisasi *centroid*. Proses *Clustering* dijalankan berulang menggunakan *K-Means++* dengan $n_{\text{init}} = 50$, kemudian konfigurasi terbaik dipilih berdasarkan nilai WCSS minimum yang konsisten. Stabilitas dievaluasi melalui konsistensi komposisi anggota dan karakteristik *centroid* pada nilai k optimal, sehingga struktur *cluster* mencerminkan pola intrinsik data.

2.11. Analisis Dinamika Spasial-Temporal Cluster

Setelah konfigurasi stabil diperoleh, analisis dilanjutkan pada dinamika keanggotaan *cluster* antar kuartal. Unit analisis $\text{State} \times \text{Quarter}$ memungkinkan identifikasi mobilitas performa wilayah secara periodik. Perubahan label *cluster* diinterpretasikan sebagai pergeseran posisi relatif performa terhadap struktur nasional. Integrasi evaluasi struktural dengan analisis temporal menghasilkan segmentasi yang tidak hanya statis, tetapi juga menggambarkan evolusi *spatio-temporal* performa distribusi secara kuartalan.

3. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian berdasarkan tahapan metodologi yang telah dirancang, mulai dari transformasi dan *preprocessing* data hingga proses *Clustering*, evaluasi, dan analisis dinamika *spatio-temporal*. Penyajian hasil difokuskan pada keterkaitan antara prosedur analisis yang digunakan dengan struktur segmentasi yang dihasilkan, sehingga interpretasi yang dilakukan merefleksikan proses metodologis secara konsisten.

3.1 Deskripsi Dataset

Dataset yang digunakan merupakan transaksi penjualan produk Adidas di Amerika Serikat pada periode 2020-2021. Data awal berbentuk *transaction-level* data, di mana setiap baris mencatat satu transaksi lengkap dengan 13 atribut, yaitu: *Retailer*, *Retailer Id*, *Invoice Date*, *Region*, *State*, *City*, *Product*, *Price per Unit*, *Units Sold*, *Total Sales*, *Operating Profit*, *Operating Margin*, dan *Sales Method*. Untuk memberikan gambaran struktur data mentah yang digunakan, 3 observasi pertama ditampilkan pada Tabel 1.

Struktur tersebut menunjukkan data masih bersifat granular pada tingkat transaksi individu, sedangkan metodologi penelitian menggunakan segmentasi wilayah berbasis *K-Means* yang memerlukan unit analisis agregat. Oleh karena itu, dilakukan transformasi unit analisis agar selaras dengan tujuan segmentasi performa wilayah berbasis dimensi spasial dan temporal. Transformasi dilakukan melalui agregasi berdasarkan variabel spasial dan temporal untuk menghasilkan indikator kuantitatif yang digunakan dalam *Clustering*, yaitu *Total Sales*, *Units Sold*, *Operating Profit*, dan *Operating Margin*. Proses ini mereduksi kompleksitas data transaksi sekaligus mempertahankan informasi utama performa distribusi tiap wilayah, sehingga *Dataset* sesuai dengan kebutuhan *K-Means* yang berbasis variabel numerik.

3.2 Transformasi Unit Analisis dan Karakteristik Data Hasil Agregasi

Transformasi unit analisis dilakukan untuk mengubah data transaksi yang bersifat granular menjadi data agregat sesuai kebutuhan Algoritma *Clustering*. Unit observasi dibentuk untuk merepresentasikan performa distribusi berdasarkan wilayah dan waktu. Agregasi dilakukan pada dimensi spasial dan temporal menggunakan kombinasi $\text{State} \times \text{Quarter}$, sehingga setiap observasi mencerminkan karakteristik penjualan suatu wilayah pada periode tertentu. Variabel agregasi terdiri dari *Total Sales*, *Units Sold*, *Operating Profit*, dan *Operating Margin* yang masing-masing merepresentasikan volume penjualan, nilai pendapatan,

profitabilitas, dan efisiensi operasional. Seluruh variabel bersifat numerik sehingga sesuai dengan algoritma K-Means yang berbasis jarak dalam mengukur kemiripan antar objek

Proses agregasi dilakukan dengan menjumlahkan *Total Sales*, *Units Sold*, dan *Operating Profit*, serta menghitung rata-rata *Operating Margin* untuk setiap kombinasi *state* dan kuartal. Hasil transformasi menghasilkan *Dataset* yang lebih ringkas, di mana jumlah observasi berkurang dari 400 menjadi 181 karena beberapa transaksi dengan kombinasi *state* dan kuartal yang sama digabungkan menjadi satu representasi. Dengan demikian, setiap observasi merepresentasikan ringkasan performa penjualan wilayah pada periode tertentu, mengurangi redundansi data sekaligus mempertahankan informasi utama. *Dataset* hasil agregasi ini memenuhi kebutuhan metodologi *Clustering* yang memerlukan data numerik terstruktur sebagai input. Untuk memberikan gambaran karakteristik data setelah agregasi, statistik deskriptif *Dataset* ditampilkan pada Tabel 2.

Tabel 1. Cuplikan Data Awal Penjualan Produk Adidas

Retailer	ID	Date	Reg		Price per Unit	Units Solds	Total Sales	Opt. Profit	Opt. Margin	Sales Method
Walmart West Gear Foot Locker	1197831	10/26/2020	South		\$45.00	950	\$427,500	\$300,000	50%	In-store
	1128299	2/17/2021	South		\$44.00	73	\$3,122	\$1,767	55%	Online
	1185732	1/5/2020	Northeast		\$60.00	900	\$540,000	\$162,000	30%	In-store

Tabel 2. Statistik Deskriptif Data Hasil Agregasi (*State* × *Quarter*)

	Total Sales (USD)	Unit Sold	Operating Profit (USD)	Operating Margin (%)
count	181	181	181	181
mean	4.971.835	13.695	1.834.999	42,04%
std	4.445.863	11.971	1.619.962	4.03%
min	138,363	284	55.749	30,25%
25%	2.401.226	7.615	898.517	39,91%
50%	3.503.123	10.577	1.316.862	42,30%
75%	5.734.880	14.363	2.166.673	44,81%
max	29.505.530	76.436	8.900.790	55,62%

Dataset hasil agregasi pada Tabel 2 terdiri dari 181 observasi yang merepresentasikan kombinasi wilayah dan periode waktu. Nilai rata-rata *Total Sales* sebesar 4,971,835 USD dengan standar deviasi 4,445,863 menunjukkan variasi performa penjualan antarwilayah yang signifikan. Variasi serupa terlihat pada *Units Sold* dan *Operating Profit* yang memiliki rentang nilai lebar, sedangkan *Operating Margin* relatif lebih stabil dengan rata-rata 42.04%, menunjukkan tingkat profitabilitas yang konsisten antarwilayah. Variasi antarvariabel ini penting bagi proses *Clustering* karena memungkinkan *K-Means* mengidentifikasi kelompok wilayah berdasarkan perbedaan karakteristik performa distribusi. Struktur data hasil agregasi juga telah sesuai dengan kebutuhan *K-Means* yang mengandalkan variabel numerik sebagai dasar pengukuran kemiripan, sehingga *Dataset* siap digunakan untuk evaluasi jumlah *cluster* dan segmentasi wilayah.

Tahapan *preprocessing* turut meningkatkan kualitas data sebelum *Clustering*. Deteksi *outlier* berbasis IQR yang diikuti winsorizing menghasilkan distribusi lebih stabil tanpa menghilangkan observasi. Selanjutnya, normalisasi Min-Max menyamakan rentang variabel sehingga mengurangi dominasi skala tertentu dalam perhitungan jarak. Dengan demikian, *Dataset* telah memenuhi asumsi algoritma *K-Means* yang sensitif terhadap perbedaan magnitudo antarvariabel.

3.3 Penentuan Jumlah Cluster Optimal

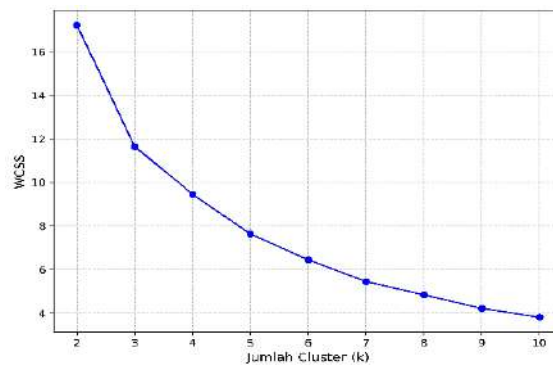
Penentuan jumlah *cluster* dilakukan dengan mengevaluasi nilai *k* pada rentang 2–10 menggunakan pendekatan multi-indeks validasi internal yang mencakup *Within-Cluster Sum of Squares* (WCSS), *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Calinski-Harabasz Index* (CHI). Evaluasi ini menghasilkan konfigurasi yang optimal secara matematis sekaligus representatif dalam menangkap variasi performa distribusi *State* × *Quarter*. Ringkasan hasil evaluasi disajikan pada Tabel 3.

Tabel 3. Hasil Evaluasi Multi-Indeks Penentuan Jumlah Cluster

k	WCSS	Silhouette	DBI	CHI
2	17.2245	0.5146	0.7803	229.5513
3	11.6235	0.4263	0.8693	211.7863
4	9.456	0.4071	0.9383	186.2929
5	7.6397	0.3435	0.9581	182.4216

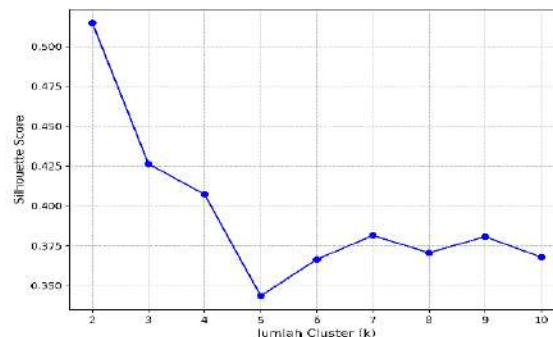
k	WCSS	Silhoute	DBI	CHI
6	6.4448	0.3663	0.9033	178.5001
7	5.4475	0.3814	0.8279	180.2862
8	4.8346	0.3705	0.8591	176.254
9	4.2129	0.3806	0.8208	179.1319
10	3.803	0.3677	0.831	177.4093

Berdasarkan Tabel 3, nilai WCSS menunjukkan penurunan paling signifikan dari $k = 2$ ke $k = 3$ dengan reduksi sebesar 32,5%, sedangkan setelah $k \geq 4$ penurunan cenderung melandai. Pola ini mengindikasikan adanya titik siku (*elbow*) pada $k = 3$, yang merepresentasikan kompromi antara kompleksitas model dan reduksi variansi *intra-cluster*. Visualisasi pola tersebut ditunjukkan pada Gambar 2.



Gambar 2. Kurva *Elbow* (WCSS terhadap Jumlah Cluster)

Evaluasi menggunakan *Silhouette Score* menunjukkan nilai tertinggi pada $k = 2$ (0,5146), sedangkan pada $k = 3$ nilai sebesar 0,4263 masih berada pada kategori moderat dan menunjukkan keseimbangan kohesi-separasi yang memadai. Penurunan nilai setelah $k \geq 4$ memperlihatkan degradasi kualitas pemisahan *cluster*. Pola perubahan nilai *Silhouette* divisualisasikan pada Gambar 3.



Gambar 3. Kurva *Silhouette Score*

Secara numerik, DBI dan CHI menunjukkan performa terbaik pada $k = 2$, namun konfigurasi dua *cluster* menghasilkan segmentasi biner yang mengompresi variasi performa wilayah menjadi dua kutub ekstrem. Dalam analisis *spasio-temporal* berbasis *State* $\times$ *Quarter*, struktur ini berpotensi menyederhanakan dinamika performa dan membatasi identifikasi wilayah transisional antar kuartal. Sebaliknya, $k = 3$ memberikan representasi yang lebih granular tanpa mengorbankan kualitas kohesi-separasi secara signifikan. Pemilihan $k = 3$ didasarkan pada *trade-off* antara optimalitas statistik dan kapasitas representasional dalam menangkap heterogenitas *spasio-temporal*. Dengan mempertimbangkan indikasi *elbow* pada WCSS, nilai *Silhouette* yang memadai, serta konsistensi antar indeks, $k = 3$ ditetapkan sebagai jumlah *cluster* paling representatif untuk analisis lanjutan.

3.4 Hasil Segmentasi *K-Means*

Berdasarkan hasil evaluasi jumlah *cluster* optimal pada tahap sebelumnya, ditetapkan $k = 3$ sebagai konfigurasi yang digunakan dalam proses segmentasi. Implementasi algoritma *K-Means* dilakukan sesuai dengan rancangan metodologi, yaitu menggunakan inisialisasi *K-Means++* dan parameter *n_init* = 50 untuk meningkatkan kualitas serta konsistensi solusi yang dihasilkan. Dengan konfigurasi tersebut, proses *Clustering* menghasilkan struktur partisi yang merupakan hasil optimasi terhadap variasi *intra-cluster*

sebagaimana dirumuskan dalam fungsi objektif *K-Means*. Distribusi hasil segmentasi terhadap 181 unit analisis (*State × Quarter*) disajikan pada Tabel 4.

Tabel 4. Distribusi Observasi per *Cluster*

<i>Cluster</i>	Jumlah Observasi	Persentase
C0	58	32,0%
C1	101	55,8%
C2	22	12,2%
Total	181	100%

Sebagian besar observasi (55,8%) berada pada *cluster* C1. *Cluster* C0 mencakup 32,0% observasi, sedangkan *cluster* C2 terdiri dari 12,2% observasi. Komposisi ini menunjukkan bahwa mayoritas wilayah berada pada kategori performa rendah-menengah, sementara wilayah dengan performa tinggi membentuk proporsi yang relatif kecil. Karakteristik rata-rata setiap variabel pada masing-masing *cluster* disajikan pada Tabel 5.

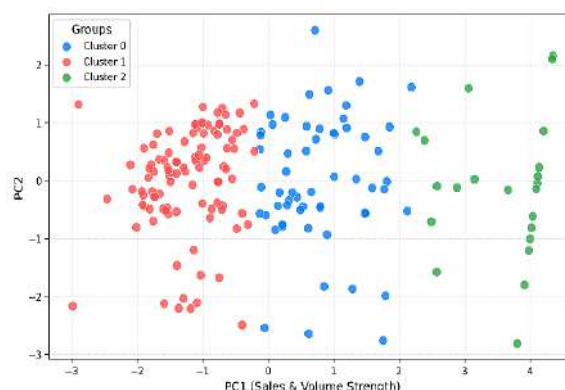
Tabel 5. Profil Rata-Rata Variabel per *Cluster*

<i>Cluster</i>	Total Sales (USD)	Unit Sold	Operating Profit (USD)	Operating Margin (%)
C0	5.626.050	14.260	2.084.473	42%
C1	2.537.176	7.955	918.685	42%
C2	9.859.056	22.780	3.764.511	43%

Perbedaan antarkluster terutama dipengaruhi oleh skala volume dan profit absolut. *Cluster* C2 menunjukkan rata-rata total sales sebesar USD 9.859.056 dan operating profit sebesar USD 3.764.511, secara signifikan lebih tinggi dibandingkan C1 yang mencatat total sales USD 2.537.176 dan operating profit USD 918.685. Jumlah unit terjual pada C2 (22.780 unit) hampir tiga kali lipat dibandingkan dengan C1 (7.955 unit). *Cluster* C0 berada pada posisi intermediari dengan total sales USD 5.626.050 dan operating profit USD 2.084.473. Sementara itu, nilai operating margin relatif homogen pada seluruh *cluster*, berada pada kisaran 0,42–0,43, yang menunjukkan bahwa perbedaan segmentasi tidak dipengaruhi oleh efisiensi operasional, melainkan oleh variasi kapasitas pasar dalam hal volume distribusi dan profit absolut.

Pola perbedaan antarkluster tersebut sejalan dengan karakteristik algoritma *K-Means* yang mengelompokkan data berdasarkan kedekatan dalam ruang fitur numerik. Karena seluruh variabel telah melalui proses normalisasi Min-Max, kontribusi setiap variabel dalam perhitungan jarak Euclidean menjadi proporsional. Dalam konteks ini, dominasi variabel Total Sales dan Operating Profit dalam membedakan *cluster* mengindikasikan bahwa kedua variabel tersebut memiliki tingkat variasi yang lebih tinggi dibandingkan variabel lainnya, sehingga berperan lebih signifikan dalam pembentukan centroid dan struktur segmentasi akhir.

Validasi separasi *cluster* dilakukan melalui reduksi dimensi menggunakan Principal Component Analysis (PCA). Variabel yang telah distandarisasi diproyeksikan ke dua komponen utama pertama untuk mengevaluasi struktur geometris *cluster* dalam ruang berdimensi rendah. Hasil proyeksi ditampilkan pada Gambar 4.



Gambar 4. *Cluster Visualization using PCA*

Visualisasi menunjukkan pemisahan *cluster* dominan terjadi pada sumbu komponen utama pertama (PC1) yang merepresentasikan kombinasi linier variabel volume dan profit. *Cluster* C2 terkonsentrasi pada nilai PC1 lebih tinggi, C1 pada rentang lebih rendah, dan C0 berada di antaranya. Pola ini konsisten dengan centroid pada Tabel 5 dan menunjukkan separasi geometris yang jelas dengan tumpang tindih terbatas. Hasil

segmentasi *K-Means* tidak hanya menggambarkan pengelompokan deskriptif, tetapi juga struktur data hasil optimasi metodologis. Struktur tersebut menunjukkan diferensiasi kapasitas distribusi nasional yang bersifat hierarkis, di mana sebagian kecil wilayah memiliki kontribusi volume dan profit jauh lebih besar dibandingkan mayoritas wilayah lainnya. Dengan demikian, segmentasi memberikan dasar empiris yang kuat untuk analisis performa wilayah dalam dinamika distribusi $State \times Quarter$.

3.5 Uji Stabilitas Cluster

Untuk memastikan hasil segmentasi tidak bergantung pada inisialisasi centroid, dilakukan uji stabilitas sebagai metode validasi tambahan dengan menjalankan kembali algoritma *K-Means* ($k = 3$) menggunakan beberapa nilai *random_state* (10, 20, 30, 40, dan 50). Konsistensi struktur *cluster* dievaluasi menggunakan Adjusted Rand Index (ARI) yang mengukur kesesuaian dua partisi tanpa memperhatikan penomoran label. Hasil menunjukkan nilai ARI sebesar 1,000 pada seluruh percobaan, di mana perbedaan label hanya berupa label permutation tanpa mengubah komposisi anggota *cluster*. Temuan ini menunjukkan struktur partisi identik pada setiap inisialisasi, sehingga konfigurasi $k = 3$ tidak sensitif terhadap variasi titik awal centroid. Dengan demikian, penggunaan ARI sebagai validasi tambahan di luar metrik internal (Silhouette, DBI, dan CHI) mengonfirmasi bahwa segmentasi merepresentasikan pola intrinsik data. Hasil ini memperkuat validitas struktural model dan menunjukkan bahwa konfigurasi *cluster* bersifat robust serta reliabel untuk analisis dinamika spasio-temporal selanjutnya.

3.6 Dinamika Spasial-Temporal Segmentasi $State \times Quarter$

Analisis lanjutan mengevaluasi dinamika keanggotaan *cluster* pada tingkat *state* selama delapan kuartal (2020-2021). Pendekatan $State \times Quarter$ mampu menangkap perubahan posisi relatif performa wilayah dalam jangka pendek. Dari 50 *state*, sebanyak 26 *state* (52%) mengalami perubahan *cluster* setidaknya satu kali, sedangkan 24 *state* (48%) konsisten pada *cluster* yang sama. Distribusi perpindahan disajikan pada Tabel 6.

Tabel 6. Distribusi Perpindahan *Cluster* per *State*

Kategori Perpindahan	Jumlah State	Persentase
Tidak berpindah <i>cluster</i>	24	48%
Berpindah antar 2 <i>cluster</i>	24	48%
Berpindah antar 3 <i>cluster</i>	2	4%
Total	50	100%

Hasil pada Tabel 6 menunjukkan bahwa hampir separuh wilayah mengalami mobilitas antardua *cluster* berbeda, dan sebagian kecil (4%) bahkan berpindah pada tiga *cluster* yang berbeda selama periode pengamatan. Temuan ini mengindikasikan bahwa struktur performa distribusi bersifat relatif dan kontekstual terhadap kondisi sistem distribusi nasional pada periode tertentu, bukan tipologi yang permanen. Untuk memberikan pemahaman yang lebih komprehensif terhadap dimensi spasial, distribusi *cluster* divisualisasikan dalam bentuk peta wilayah pada dua titik waktu berbeda, yaitu awal periode observasi (2020Q1) dan akhir tahun kedua (2021Q4), sebagaimana ditunjukkan pada Gambar 5 dan Gambar 6.



Gambar 5. Peta Spasial Distribusi *Cluster* per *State* (2020Q1)

Perbandingan visual antara Gambar 5 dan Gambar 6 menunjukkan perubahan distribusi *cluster* antar wilayah dalam periode singkat. Pada 2020Q1, *cluster* masih terbatas dan terkonsentrasi pada sedikit wilayah, mencerminkan fase awal segmentasi. Pada 2021Q4, jumlah wilayah yang tersegmentasi meningkat dengan penyebaran geografis yang lebih luas. Perubahan ini menunjukkan struktur performa distribusi bersifat dinamis dan berkembang secara spasial. Beberapa wilayah bergeser antar *cluster*, mencerminkan perubahan kapasitas distribusi dan dinamika pasar regional. Visualisasi ini menegaskan bahwa segmentasi merepresentasikan posisi kompetitif wilayah yang dapat berubah seiring perkembangan kondisi pasar dan

distribusi. Untuk melengkapi dimensi temporal, distribusi observasi tiap *cluster* per kuartal disajikan pada Tabel 7.

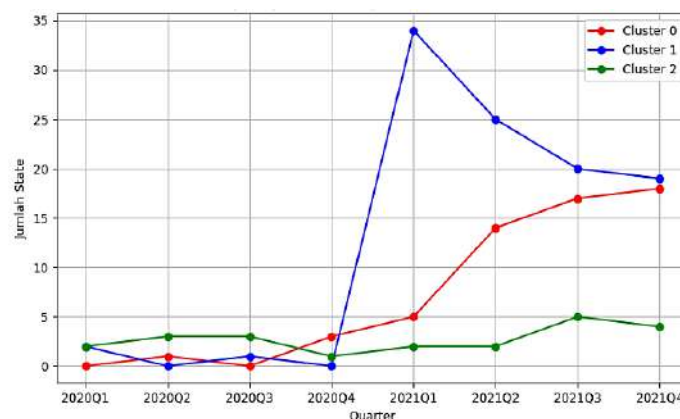


Gambar 6. Peta Spasial Distribusi *Cluster* per State (2021Q4)

Tabel 7. Distribusi Observasi per *Cluster* pada Setiap Quarter

Quarter	C0	C1	C2
2020Q1	0	2	2
2020Q2	1	0	3
2020Q3	0	1	3
2020Q4	3	0	1
2021Q1	5	34	2
2021Q2	14	25	2
2021Q3	17	20	5
2021Q4	18	19	4

Untuk memperjelas pola perubahan distribusi *cluster* antar periode, dilakukan visualisasi temporal terhadap jumlah observasi pada masing-masing *cluster* di setiap kuartal sebagaimana ditunjukkan pada Gambar 7.



Gambar 7. Grafik Temporal Distribusi *Cluster* per Quarter

Visualisasi pada Gambar 7 menunjukkan dinamika yang lebih jelas dibandingkan tabel. Pada tahun 2020, distribusi *cluster* masih terbatas dan fluktuatif dengan jumlah observasi kecil. Memasuki 2021, jumlah observasi meningkat signifikan, terutama pada *cluster* performa rendah (C1) yang mencapai puncak pada 2021Q1 sebelum menurun bertahap, menunjukkan sebagian besar wilayah awalnya berada pada performa rendah sebelum mengalami redistribusi. *Cluster* menengah (C0) menunjukkan tren peningkatan konsisten sepanjang 2021, mengindikasikan pergeseran wilayah dari performa rendah menuju kondisi lebih stabil. *Cluster* performa tinggi (C2) tetap pada proporsi kecil dan relatif stabil dengan sedikit peningkatan pada kuartal tertentu. Pola ini menunjukkan dinamika performa distribusi terjadi secara spasial dan temporal, mencerminkan penyesuaian kapasitas distribusi serta perubahan kondisi pasar secara bertahap.

Tabel 7 menunjukkan struktur *cluster* tahun 2020 masih terbatas, sedangkan tahun 2021 observasi meningkat dan distribusi lebih seimbang, terutama melalui kenaikan C0 dan redistribusi *cluster* lain. Pergeseran ini mencerminkan restrukturisasi performa distribusi nasional dan perubahan posisi relatif wilayah. Dinamika tersebut tidak bertentangan dengan uji stabilitas ($ARI = 1,000$) karena stabilitas menunjukkan konsistensi struktur model, sementara mobilitas pada 52% *state* merepresentasikan perubahan intrinsik antar kuartal akibat penggunaan unit analisis *State* $\times$ *Quarter*. Integrasi dimensi spasial dan temporal menghasilkan segmentasi yang lebih sensitif dibanding agregasi tahunan yang cenderung mereduksi

dinamika jangka pendek. Pendekatan ini tidak hanya memetakan tipologi performa wilayah, tetapi juga mengungkap mobilitas dan pergeseran posisi relatif sebagai dasar evaluasi distribusi nasional yang lebih adaptif dan kontekstual.

3.7 Interpretasi Manajerial

Hasil segmentasi *K-Means* ($k = 3$) mengidentifikasi tiga tingkat performa distribusi: C2 (12,2%) sebagai performa tinggi, C0 (32,0%) menengah, dan C1 (55,8%) rendah. *Cluster* C2 memiliki rata-rata *Total Sales* USD 9.859.056, *Units Sold* 22.780, dan *Operating Profit* USD 3.764.511 sehingga menjadi kontributor utama; strategi difokuskan pada mempertahankan kinerja, optimalisasi stok, dan benchmark praktik terbaik. *Cluster* C0 dengan *Total Sales* USD 5.626.050 dan *Operating Profit* USD 2.084.473 memiliki potensi peningkatan, sehingga diarahkan pada penguatan penetrasi pasar dan optimalisasi channel distribusi. *Cluster* C1 mencakup mayoritas observasi dengan *Total Sales* USD 2.537.176 dan *Operating Profit* USD 918.685; strategi difokuskan pada evaluasi efisiensi distribusi dan intervensi selektif berdasarkan potensi wilayah. Analisis dinamika menunjukkan 52% *state* berpindah *cluster* antar kuartal, menandakan performa wilayah tidak permanen. Wilayah berperforma rendah dapat bertransisi ke kategori lebih tinggi, sehingga strategi distribusi perlu adaptif dan dievaluasi berkala. Prioritas pengembangan difokuskan pada wilayah yang bergerak dari C1 ke C0 atau C2, memungkinkan alokasi sumber daya lebih efisien sekaligus menjaga stabilitas wilayah unggulan.

3.8 Diskusi

Hasil penelitian ini sejalan dengan studi sebelumnya yang menggunakan algoritma *K-Means* untuk segmentasi performa penjualan atau wilayah. Namun, penelitian terdahulu umumnya menggunakan agregasi tahunan atau wilayah statis sehingga belum mampu menangkap dinamika perubahan performa secara temporal. Berbeda dengan pendekatan tersebut, penelitian ini mengintegrasikan dimensi spasial dan temporal melalui unit analisis *State × Quarter*. Pendekatan ini menunjukkan bahwa 52% *state* mengalami perubahan *cluster* antar kuartal, fenomena yang tidak terdeteksi dalam agregasi tahunan. Selain itu, penelitian ini menambahkan evaluasi stabilitas menggunakan *Adjusted Rand Index* (ARI) sebagai validasi tambahan untuk mengukur konsistensi hasil *Clustering*, yang belum banyak digunakan pada penelitian segmentasi distribusi. Dengan demikian, penelitian ini tidak hanya melakukan segmentasi wilayah, tetapi juga mengungkap dinamika perubahan performa secara periodik dengan evaluasi kualitas dan stabilitas *cluster* yang komprehensif, sehingga memberikan interpretasi distribusi wilayah yang lebih akurat.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model segmentasi performa distribusi berbasis integrasi *State × Quarter* menggunakan algoritma *K-Means* yang menunjukkan konfigurasi optimal dan stabil. Penentuan jumlah *cluster* $k = 3$ didukung oleh evaluasi multi-indeks (Elbow, Silhouette, Davies-Bouldin, dan Calinski-Harabasz), sementara uji stabilitas menghasilkan nilai ARI = 1,000 yang menunjukkan konsistensi struktur *cluster*. Hasil segmentasi mengelompokkan 181 observasi *spatio-temporal* ke dalam tiga kategori performa (rendah, menengah, dan tinggi) yang terutama dibedakan oleh skala total *sales* dan *operating profit*, sementara *operating margin* relatif homogen. Analisis *spatio-temporal* juga menunjukkan bahwa 52% *state* mengalami perpindahan *cluster* antarkuartal, yang menegaskan bahwa performa wilayah bersifat dinamis dan tidak sepenuhnya terwakili melalui agregasi tahunan. Temuan ini menunjukkan bahwa integrasi dimensi spasial dan temporal memberikan segmentasi yang lebih representatif dalam menangkap variasi performa distribusi antarwilayah dan antarperiode. Dengan demikian, pendekatan *State × Quarter* dapat menjadi kerangka analitis yang lebih adaptif dalam mendukung evaluasi dan perencanaan strategi distribusi berbasis data. Untuk penelitian selanjutnya, analisis dapat diperluas dengan membandingkan kinerja *K-Means* dengan algoritma *Clustering* lain seperti Hierarchical *Clustering*, DBSCAN, atau Gaussian Mixture Model untuk mengevaluasi konsistensi struktur segmentasi yang dihasilkan. Selain itu, penggunaan periode data yang lebih panjang juga disarankan agar dinamika perubahan *cluster* antarwilayah dapat diamati secara lebih komprehensif serta memungkinkan analisis tren *spatio-temporal* dalam jangka panjang.

REFERENSI

- [1] D. Abid, R. W. Adikusuma, A. M. AL Fikri, and R. K. Hapsari, "Penerapan Metode *K-Means Clustering* Untuk Analisa Penjualan Komoditas Toko Tani Indonesia," *KERNEL J. Ris. Inov. Bid. Inform. dan Pendidik. Inform.*, vol. 3, no. 2, pp. 25–30, 2023, doi: 10.31284/j.kernel.2022.v3i2.4076.
- [2] N. Afiasari, N. Suarna, and N. Rahaningsi, "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma *Clustering* dengan Metode *K-Means*," *J. SAINTEKOM*, vol. 13, no. 1, pp. 100–110, 2023, doi: 10.33020/saintekom.v13i1.402.
- [3] R. I. Manarung, E. Widodo, and A. M. Rifai, "Sales Data *Clustering* Using the *K-Means* Algorithm to Determine Retail Product Needs," vol. 5, no. April, pp. 226–234, 2025.
- [4] B. I. Nugroho, A. Rafhina, P. S. Ananda, and G. Gunawan, "Customer segmentation in sales

- transaction data using *K-Means Clustering* algorithm,” *J. Intell. Decis. Support Syst.*, vol. 7, no. 2, pp. 130–136, 2024, doi: 10.35335/idss.v7i2.236.
- [5] A. S. Harish and C. Malathy, “Customer Segment Prediction on Retail Transactional Data Using *K-Means* and Markov Model,” 2023, doi: 10.32604/iasc.2023.032030.
- [6] M. M. Ridzki, I. Hadijah, M. Mukidin, A. Azzahra, and A. Nurjanah, “*K-Means* Algorithm Method for *Clustering* Best-Selling Product Data at XYZ Grocery Stores,” *Int. J. Soc. Serv. Res.*, vol. 3, no. 12, pp. 3354–3367, 2023, doi: 10.46799/ijssr.v3i12.652.
- [7] E. F. L. Awalina and W. I. Rahayu, “Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan *K-Means Clustering* pada Transaksi Online Retail,” *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023, doi: 10.34010/jati.v13i2.10090.
- [8] Y. Putri, D. Aldo, and W. Ilham, “Retail Marketing Strategy Optimization: Customer Segmentation with Artificial Intelligence Integration and *K-Means Clustering*,” *Sinkron*, vol. 8, no. 4, pp. 2155–2163, 2024, doi: 10.33395/sinkron.v8i4.14000.
- [9] E. Omol, D. Onyangor, L. Mburu, and P. Abuonji, “Application Of *K-Means Clustering* For Customer Segmentation In Grocery Stores In Kenya,” *Int. J. Sci. Technol. Manag.*, vol. 5, no. 1, pp. 192–200, 2024, doi: 10.46729/ijstm.v5i1.1024.
- [10] J. M. John, O. Shobayo, and B. Ogunleye, “An Exploration of *Clustering* Algorithms for Customer Segmentation in the UK Retail Market,” *Analytics*, vol. 2, no. 4, pp. 809–823, 2023, doi: 10.3390/analytics2040042.
- [11] P. Anitha and M. M. Patil, “RFM model for customer purchase behavior using *K-Means* algorithm,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1785–1792, 2022, doi: 10.1016/j.jksuci.2019.12.011.
- [12] B. K. Kristanto, S. Widyayuningtias, P. Listio, M. Amien, and P. I. Baskoro, “Visit Recommendation Model : Recursive *K-Means Clustering* Analysis of Retail Sales Data,” vol. 8, no. 1, pp. 221–225, 2024.
- [13] R. Angeline, “Perancangan Segmentasi Pasar Sepatu Lokal Pillary Footwear Menggunakan *K-Means Clustering*,” vol. 12, no. 1, pp. 734–739, 2025.
- [14] S. Sinaga, R. Ananda, H. Q. Karima, and A. M. Tazuddin, “Marketing Analysis of Shoe Products Using Principal Coordinates Analysis and *K-Means Clustering* Based on the Marketing Mix at Bintang Sepatu Purwokerto MSME,” vol. 6, no. 3, pp. 1405–1418, 2025.
- [15] D. Zekri, “Spatiotemporal *Clustering* of Parking Lots at the City Level for Efficiently Sharing Occupancy Forecasting Models,” pp. 1–25, 2023.
- [16] X. He, “A *Spatio-temporal* Feature Trajectory *Clustering* Algorithm Based on Deep Learning,” 2022.
- [17] O. Dorabiala, D. V. Dabke, J. Webster, J. N. Kutz, and A. Aravkin, “Spatiotemporal *k*-means,” pp. 1–18, 2024.
- [18] Y. T. Tesfaldet, M. G. Elmubarak, and N. Al Hosani, “Extraction of Urban Quality of Life Indicators Using Remote Sensing and Machine Learning : The Case of Al Ain City , United Arab Emirates (UAE),” 2022.
- [19] H. Mulyani, R. A. Setiawan, and H. Fathi, “Optimization of *K* Value in *Clustering* Using *Silhouette Score* (Case Study: Mall Customers Data),” *J. Inf. Technol. Its Util.*, vol. 6, no. 2, pp. 45–50, 2023, doi: 10.56873/jitu.6.2.5243.
- [20] I. F. Ashari, E. D. Nugroho, R. Baraku, I. N. Yanda, and R. Liwardana, “Analysis of Elbow , *Silhouette* , *Davies-Bouldin* , *Calinski-Harabasz* , and *Rand-Index* Evaluation on *K-Means* Algorithm for Classifying Flood- Affected Areas in Jakarta,” vol. 7, no. 1, pp. 95–103, 2023.