

**ANALISIS HUBUNGAN JUMLAH PENUMPANG DAN
BUS SEKOLAH DI KOTA JAKARTA PADA TAHUN
2017 HINGGA 2019 MENGGUNAKAN PENDEKATAN
*UNSUPERVISED LEARNING***

TUGAS AKHIR

**Diajukan Untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana
Sistem Komputer**



Oleh:

AURELL OCTAVIONA SINAGA

09011382126149

JURUSAN SISTEM KOMPUTER

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2025

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS HUBUNGAN JUMLAH PENUMPANG DAN BUS SEKOLAH DI KOTA JAKARTA PADA TAHUN 2017 HINGGA 2019 MENGGUNAKAN PENDEKATAN UNSUPERVISED LEARNING

Sebagai salah satu syarat untuk penyelesaian studi di Program Studi S1 Sistem
Komputer

Oleh:
AURELL OCTAVIONA SINAGA
09011382126149

Pembimbing 1 : **Dr. Rossi Passarella, S.T., M.Eng.**
NIP. 197806112010121004

Mengetahui
Ketua Jurusan Sistem Komputer



Dr. Ir. Sukemi, M.T
196612032006041001

AUTHENTICATION PAGE

FINAL TASK

ANALYZING THE RELATIONSHIP BETWEEN THE NUMBER OF PASSENGERS AND SCHOOL BUSES IN THE CITY OF JAKARTA IN 2017-2019 USING AN UNSUPERVISED LEARNING APPROACH

As one of the requirements for the completion of the Bachelor's Degree Program
in Computer Systems

By:

AURELL OCTAVIONA SINAGA
09011382126149

Supervisor 1

: **Dr. Rossi Passarella, S.T., M.Eng.**
NIP. 197806112010121004

Mengetahui

Ketua Jurusan Sistem Komputer



Dr. Ir. Sukemi, M.T
196612032006041001

HALAMAN PERSETUJUAN

Pada hari jumat tanggal 23 Mei 2025 telah dilaksanakan ujian sidang skripsi oleh
Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Aurell Octaviona Sinaga

NIM : 09011382126149

Judul : Analisis Hubungan Jumlah Penumpang dan Bus Sekolah di Kota
Jakarta pada tahun 2017 hingga 2019 menggunakan pendekatan
Unsupervised Learning

Tim Penguji

1. Ketua : Huda Ubaya, S.T., M.T.

2. Penguji : Sutarno S.T., M.T.

3. Pembimbing : Dr. Rossi Passarella, S.T., M.Eng.

Mengetahui, 12/6/25
Ketua Jurusan Sistem Komputer,



Dr. Ir. Sukemi, M.T.

NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Aurell Octaviona Sinaga
NIM : 09011382126149
Program Studi : Sistem Komputer
Judul : Analisis Hubungan Jumlah Penumpang dan Bus Sekolah di Kota Jakarta pada tahun 2017 hingga 2019 menggunakan pendekatan *Unsupervised Learning*

Hasil Pengecekan Plagiat/Turnitin : 3%

Menyatakan bahwa laporan tugas akhir ini adalah hasil karya saya sendiri dan tidak mengandung unsur penjiplakan atau plagiarisme. Apabila ditemukan unsur plagiarisme dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku. Demikian, pernyataan ini saya buat dengan kesadaran penuh dan tanpa adanya unsur paksaan dari pihak manapun.



Indralaya, Juni 2025



Aurell Sinaga

KATA PENGANTAR

Segala puji dan rasa syukur penulis panjatkan pada Tuhan Yang Maha Esa, atas segala karunia dan berkat-Nya penulis dapat menyelesaikan tugas akhir yang berjudul “Analisis Hubungan Jumlah Penumpang dan Bus Sekolah di kota Jakarta pada tahun 2017 hingga 2019 menggunakan pendekatan *Unsupervised Learning*” dengan lancar dan baik guna untuk menyelesaikan Program Studi Sistem Komputer di Universitas Sriwijaya.

Penulis memilih topik ini karena pentingnya memahami pola penggunaan transportasi sekolah di kota Jakarta, khususnya dalam konteks jumlah dan jumlah penumpang serta potensi penerapan teknik machine learning, khususnya unsupervised learning dalam menganalisis data tersebut secara mendalam tanpa label. Penulis juga berharap agar tulisan ini dapat memberikan manfaat maupun sebagai sumber untuk penelitian lainnya.

Melalui kesempatan ini, penulis ingin mengucapkan terimakasih kepada semua pihak yang mendukung dan terlibat dalam proses yang panjang ini. Penulis mengucapkan terimakasih kepada:

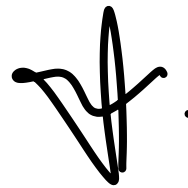
1. Tuhan Yang Maha Esa, yang telah melimpahkan berkat dan karunia Nya sehingga penulis dapat menyelesaikan tugas akhir ini dengan lancar dan tepat waktu.
2. Kedua orang tua yaitu Bapak Jamson Sinaga dan Ibu Enita yang saya cintai dan saya banggakan, serta adik saya Marcell Sinaga yang terkasih. Karena tanpa dukungan dan doa kalian, penulis tidak akan bisa mencapai tahap ini.
3. Bapak Dr. Ir. Sukemi, M.T., selaku Ketua Jurusan Sistem Komputer Universitas Sriwijaya.
4. Bapak Huda Ubaya, M.T., selaku Sekretaris Jurusan Sistem Komputer Universitas Sriwijaya.
5. Bapak Dr. Rossi Passarella, M.Eng., selaku Dosen Pembimbing Tugas Akhir yang telah memberikan arahan, bimbingan serta masukan yang sangat berarti dalam penyusunan tugas akhir ini.

6. Bapak Muhammad Ali Buchari, M.T., selaku Dosen Penasehat Akademik Jurusan Sistem Komputer, Universitas Sriwijaya, atas dukungan dan arahnya untuk bisa lulus tepat waktu.
7. Saya juga memberikan rasa terimakasih untuk teman-teman saya, terkhususnya Eva Rianty dan Jessica yang selalu menjadi tempat berkeluh kesah penulis, dan selalu memberikan dukungan baik secara tindakan maupun perkataan.
8. Dan yang terakhir, saya ucapkan rasa syukur dan bangga pada diri sendiri karena telah berhasil melewati segala proses perkuliahan dengan sabar dan tekun terhadap semua hal dan masalah yang telah terjadi.

Penulis menyadari bahwa penulisan ini masih jauh dari kata sempurna, baik dari segi isi dan penyajiannya. Oleh karena itu penulis dengan sangat terbuka terhadap kritik dan saran yang membangun untuk kedepannya. Maka dari itu penulis berharap, dengan adanya penulisan tugas akhir ini menjadi manfaat bagi pembaca dan penelitian sejenis.

Palembang, Juni 2025

Penulis,



Aurell Sinaga

NIM. 09011382126149

**ANALISIS HUBUNGAN JUMLAH PENUMPANG DAN BUS SEKOLAH
DI KOTA JAKARTA PADA TAHUN 2017 HINGGA 2019
MENGUNAKAN PENDEKATAN UNSUPERVISED LEARNING**

AURELL OCTAVIONA SINAGA (09011382126149)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email: vionaaurellie@gmail.com

ABSTRAK

Penelitian ini bertujuan mengidentifikasi pola hubungan antara jumlah bus dan jumlah penumpang sekolah di Jakarta pada periode 2017–2019 menggunakan pendekatan *unsupervised learning*. Empat algoritma *clustering* yaitu K-Means, Fuzzy C-Means, Gaussian Mixture Model, dan Spectral Clustering digunakan untuk mengelompokkan data, dan semuanya menghasilkan dua klaster optimal, namun dari keempat algoritma tersebut, K-Means menjadi algoritma paling unggul berdasarkan hasil metrik evaluasi dan dipilih sebagai metode untuk menganalisis hasil klaster. Klaster pertama menunjukkan rasio bus terhadap penumpang yang tinggi, sedangkan klaster kedua menunjukkan rasio yang rendah. Pola ini menjadi dasar untuk perencanaan transportasi sekolah yang lebih efisien. Hasil *clustering* selanjutnya digunakan sebagai label untuk melatih model *Decision Tree Classifier*, yang kemudian diaplikasikan pada data tahun 2021. Model ini berhasil mengklasifikasikan data baru ke dalam dua klaster sesuai pola yang telah terbentuk sebelumnya.

Kata Kunci: Unsupervised Learning, Clustering, K-Means, Fuzzy C-Means, Gaussian Mixture Model, Spectral Clustering

**ANALYZING THE RELATIONSHIP BETWEEN THE NUMBER OF
PASSENGERS AND SCHOOL BUSES IN THE CITY OF JAKARTA IN
2017-2019 USING AN UNSUPERVISED LEARNING APPROACH.**

AURELL OCTAVIONA SINAGA (09011382126149)

*Computer Engineering Department, Faculty of Computer Science, Sriwijaya
University*

Email : vionaaurellie@gmail.com

ABSTRACT

This study aims to identify patterns in the relationship between the number of school buses and the number of school passengers in Jakarta during the 2017–2019 period using an unsupervised learning approach. Four clustering algorithms K-Means, Fuzzy C-Means, Gaussian Mixture Model, and Spectral Clustering were applied to group the data, all of which produced two optimal clusters. Among these algorithms, K-Means outperformed the others based on evaluation metrics and was selected as the primary method for analyzing the clustering results. The first cluster represents a high ratio of buses to passengers, while the second shows a lower ratio. These patterns serve as a foundation for more efficient school transportation planning. The clustering results were then used as labels to train a Decision Tree Classifier model, which was applied to data from 2021. The model successfully classified the new data into the two clusters consistent with the previously identified patterns.

Keywords: Unsupervised Learning, Clustering, K-Means, Fuzzy C-Means, Gaussian Mixture Model, Spectral Clustering

DAFTAR ISI

HALAMAN PENGESAHAN.....	i
AUTHENTICATION PAGE.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PERNYATAAN	iv
KATA PENGANTAR.....	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Perumusan dan Batasan Masalah.....	2
1.2.1. Rumusan Masalah.....	3
1.2.2. Batasan Masalah	3
1.3 Tujuan dan Manfaat	4
1.3.1. Tujuan Penelitian	4
1.3.2. Manfaat Penelitian	4
1.4 Metode Penelitian	5
1.4.1. Metode Studi Pustaka dan Literatur.....	5
1.4.2. Metode Pengujian Model	5
1.4.3. Metode Analisis dan Kesimpulan	6
1.5 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	8
2.1 Machine Learning (Pembelajaran Mesin)	8
2.2 Exploratory Data Analysis (EDA)	9
2.2.1. Label Encoding	10
2.2.2. Matriks Korelasi dan Heatmap	10
2.2.3. Analisis Data Eksploratif	11

2.3 Unsupervised Learning	12
2.4 Model Clustering	13
2.4.1. K-Means.....	14
2.4.2. Fuzzy C-Means (FCM)	15
2.4.3. Gaussian Mixture Model (GMM)	16
2.4.4. Spectral Clustering	18
2.5 Matrik Evaluasi Klasterisasi	21
2.5.1. Silhouette Score	22
2.5.2. Davies-Bouldin Index (DBI).....	23
2.5.3. Calinski-Harabasz Index (CHI)	25
2.6 Analisis Hasil Klaster Optimal	26
2.6.1. Penghitungan Rata-rata	26
2.6.2. Standar Deviasi, Minimum dan Maksimum tiap fitur	27
2.6.3. T-test.....	28
2.7 Decision Tree Classifier.....	28
2.8 Visualisasi dalam model clustering	29
2.8.1. Scatter Plot	29
2.8.2. Boxplot.....	30
2.8.3. Histogram.....	31
BAB III METODE PENELITIAN.....	33
3.1 Dataset	33
3.2 Kerangka Kerja Penelitian	36
3.2.1. Data Preparation (Persiapan Dataset)	36
3.2.2. Pencarian Klaster Optimal dan Evaluasi Model Klasterisasi	39
3.2.3. Analisis Hasil Clustering	45
3.2.4. Model Prediksi	46
BAB IV HASIL DAN PEMBAHASAN.....	47
4.1 Evaluasi dan Pemilihan Model Clustering	47
4.2 Analisis Hasil Klaster	48
4.2.1. Analisis statistik tiap klaster	48
4.2.2. Analisis Jumlah Klaster tiap tahun	53

4.2.3. Interpretasi Karakteristik tiap klaster	54
4.3 Model Prediksi untuk tahun mendatang	57
4.4 Rekomendasi dari Hasil Analisis	59
BAB V PENUTUP	61
5.1 Kesimpulan	61
5.2 Saran	61
DAFTAR PUSTAKA	63

DAFTAR GAMBAR

Gambar 3.1 Flowchart Dataset	33
Gambar 3. 2 Korelasi Matriks setiap Fitur	35
Gambar 3.3 Alur Penelitian	36
Gambar 3.4 Pustaka dan fungsi K-Means	37
Gambar 3.5 Pustaka dan fungsi FCM.....	37
Gambar 3.6 Pustaka dan fungsi GMM	37
Gambar 3.7 Pustaka dan fungsi Spectral Clustering	38
Gambar 3.8 Preprocessing Dataset	38
Gambar 3.9 Pencarian Klaster Optimal K-Means	39
Gambar 3.10 Diagram Batang K-Means	40
Gambar 3.11 Pencarian Klaster Optimal FCM.....	41
Gambar 3.12 Diagram Batang FCM.....	42
Gambar 3.13 Pencarian Klaster Optimal GMM	42
Gambar 3.14 Diagram Batang GMM	43
Gambar 3.15 Pencarian Klaster Optimal Spectral Clustering	44
Gambar 3.16 Diagram Batang Spectral Clustering	45
Gambar 4. 1 Boxplot jumlah bus per klaster	50
Gambar 4. 2 Boxplot jumlah penumpang per klaster	51
Gambar 4. 3 Histogram jumlah penumpang tiap klaster	52
Gambar 4. 4 Histogram jumlah bus tiap klaster	52
Gambar 4. 5 Scatter Plot K-Means	54
Gambar 4. 6 Scatter plot dengan ellipse K-Means	55
Gambar 4. 7 Confusion matrix Decision Tree	58

DAFTAR TABEL

Tabel 3. 1 Output Klaster Optimal K-Means.....	40
Tabel 3.2 Output Klaster Optimal FCM	41
Tabel 3.3 Output Klaster Optimal GMM.....	43
Tabel 3.4 Output Klaster Optimal Spectral Clustering.....	44
Tabel 4. 1 Hasil indeks pengukuran 4 model clustering.....	47
Tabel 4. 2 Rata-rata x dan y	48
Tabel 4. 3 standar deviasi, min, max per fitur di tiap klaster K-Means.....	49
Tabel 4. 4 Jumlah Klaster per tahun	53
Tabel 4. 5 T-test K-Means.....	56
Tabel 4. 6 Evaluasi model decision tree	58
Tabel 4. 7 Hasil Prediksi.....	59

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan kehidupan kota yang signifikan di Jakarta memberikan dampak terhadap meningkatnya mobilitas masyarakat. Sarana transportasi yang tersedia di kota ini dapat dikategorikan menjadi dua jenis yaitu kendaraan pribadi terdiri dari mobil dan sepeda motor, sedangkan transportasi umum terdiri dari bus, kereta rel listrik (KRL), TransJakarta, dan *Mass Rapid Transit* (MRT)[1]. Namun, pertumbuhan jumlah kendaraan pribadi yang lebih cepat dibandingkan infrastruktur jalan menyebabkan berbagai permasalahan, seperti kemacetan lalu lintas, peningkatan polusi udara, serta kemacetan lalu lintas[2]. Dalam mengatasi masalah ini, berbagai solusi telah diupayakan oleh pemerintah, salah satunya adalah meningkatkan penggunaan transportasi umum dengan menyediakan layanan yang lebih efisien, nyaman, dan terjangkau bagi masyarakat.

Salah satu bentuk transportasi umum yang disediakan untuk kelompok tertentu di Jakarta adalah bus sekolah, yang diperuntukkan bagi pelajar dan mahasiswa. Kehadiran bus sekolah bertujuan untuk mengurangi ketergantungan siswa terhadap kendaraan pribadi dan meningkatkan keselamatan serta efisiensi perjalanan mereka. Namun, dalam operasionalnya, masih terdapat tantangan dalam hal pengelolaan layanan bus sekolah, yaitu ketidakseimbangan antara jumlah bus dan jumlah penumpang yang dapat berdampak pada efisiensi layanan[3] serta anggaran yang dikeluarkan oleh pemerintah. Maka dari itu, diperlukan strategi yang tepat dalam pengelolaan transportasi sekolah guna memastikan distribusi layanan yang optimal bagi para pelajar di Jakarta.

Adapun dua penelitian terdahulu yang relevan dengan topik yang diangkat, penelitian pertama membahas tentang layanan bus sekolah di Jakarta menggunakan metode machine learning yaitu klasifikasi untuk memprediksi jenis bus yang digunakan. Hasil dari penelitian tersebut memberikan wawasan tentang berbagai faktor yang mempengaruhi klasifikasi jenis bus serta menawarkan model prediktif yang dapat membantu pengambilan keputusan dalam pengelolaan transportasi

sekolah. Namun, penelitian tersebut masih memiliki keterbatasan yaitu adanya ketidak seimbangan data, dimana distribusi data target yaitu “jenis bus” membagi kelas menjadi dua, kelas 0 menunjukkan total bus kecil sebanyak 396 dan kelas 1 menunjukkan total bus sedang sebanyak 612. Hal ini berpotensi menyebabkan bias model dalam memprediksi kelas mayoritas. Oleh karena itu, penulis memilih untuk menitikberatkan evaluasi model pada metrik F1-score dibandingkan akurasi biasa.

Berbeda dengan pendekatan klasifikasi yang digunakan dalam penelitian pertama, penelitian kedua menerapkan pendekatan unsupervised learning dengan metode clustering[4], yang memberikan perspektif baru dalam menganalisis data transportasi. Penelitian ini mengelompokkan data kecelakaan lalu lintas, dan berhasil menemukan pola-pola tersembunyi berdasarkan lokasi, waktu, dan faktor penyebab kecelakaan.

Karena adanya perbedaan fokus penelitian dan metode dari dua penelitian terdahulu, hal itu menjadi celah penelitian ini. Sehingga penelitian yang berjudul **"Analisis Hubungan Jumlah Penumpang dan Bus Sekolah di kota Jakarta pada tahun 2017 hingga 2019 menggunakan pendekatan *Unsupervised Learning*"** bertujuan untuk menemukan pola tersembunyi pada hubungan jumlah penumpang dan jumlah bus sekolah di kota Jakarta dari tahun 2017-2019 menggunakan pendekatan unsupervised learning dengan metode *clustering*. Penelitian ini melakukan uji coba terhadap empat metode clustering yaitu K-Means, Fuzzy C-Means, Gaussian Mixture Model, dan Spectral Clustering. Dimana dari keempat metode tersebut akan dipilih salah satu metode terbaik untuk menganalisis hasil *clustering*, dan hasil *clustering* tersebut akan menjadi acuan untuk memprediksi dataset di tahun berikutnya dimana model terbaik dalam memprediksi dataset tahun berikutnya adalah klasifikasi menggunakan Decision Tree Classifier.

1.2 Perumusan dan Batasan Masalah

Bagian ini akan menjabarkan tentang perumusan masalah yang membantu untuk mengidentifikasi fokus utama penelitian, sedangkan batasan masalah berfokus kepada penetapan ruang lingkup agar penelitian ini tidak melebar dan meluas dari topik yang dibahas.

1.2.1. Rumusan Masalah

Berdasarkan uraian latar belakang di atas, diperlukan analisis yang tepat untuk mengidentifikasi pola hubungan antara jumlah bus dan jumlah penumpang dengan pendekatan Unsupervised Learning, yang didefinisikan ke dalam beberapa pertanyaan di bawah ini:

1. Apa model *clustering* yang paling efektif dalam mengelompokkan data jumlah penumpang dan jumlah bus sekolah di kota Jakarta tahun 2017-2019?
2. Apa karakteristik dari masing-masing klaster yang terbentuk dalam analisis *clustering* terhadap jumlah bus dan jumlah penumpang?
3. Bagaimana perubahan distribusi klaster berdasarkan kebutuhan transportasi sekolah di kota Jakarta dari tahun 2017-2019, dan apa implikasinya terhadap peningkatan kebutuhan transportasi?
4. Apakah hasil analisis *clustering* tersebut dapat dijadikan prediksi untuk menentukan *cluster* pada data tahun berikutnya?

1.2.2. Batasan Masalah

Batasan masalah ini dibuat untuk memperjelas lingkup penelitian supaya tetap terarah dan sesuai dengan tujuan yang ditetapkan sehingga tidak dapat dipengaruhi oleh berbagai faktor eksternal. Dimana batasan-batasan yang ditetapkan adalah sebagai berikut:

1. Penelitian ini dibatasi pada analisis hubungan antara jumlah bus sekolah dengan jumlah penumpang sekolah yang merupakan pelajar di Kota Jakarta selama periode tahun 2017 hingga 2019. Data yang digunakan hanya mencakup pelajar yang tercatat menggunakan layanan bus sekolah yang disediakan oleh pemerintah daerah, dan tidak mencakup penumpang dari kalangan umum.
2. Pendekatan yang digunakan adalah unsupervised learning dengan melakukan uji coba terhadap keempat metode *clustering* yaitu K-Means, Fuzzy C-Means, GMM, dan Spectral Clustering. Dan hanya satu metode terpilih, yaitu metode dengan hasil akurasi terbaik berdasarkan evaluasi matriks guna untuk melakukan analisis hasil *clustering*.

3. Penelitian ini menggunakan bahasa Python di lingkungan Google Colab.

1.3 Tujuan dan Manfaat

Tujuan dari penelitian ini adalah untuk memberikan pemahaman lebih mendalam mengenai pola hubungan antara jumlah bus dan jumlah penumpang, dimana dari tujuan tersebut diharapkan dapat memberikan sejumlah manfaat bagi pemangku kepentingan maupun pembaca.

1.3.1. Tujuan Penelitian

Adapun tujuan dari penelitian yang dilakukan, akan dijelaskan secara rinci ke dalam poin-poin di bawah ini:

1. Untuk menentukan model clustering yang paling efektif dalam mengelompokkan data jumlah bus dan jumlah penumpang sekolah di kota Jakarta dari tahun 2017-2019.
2. Untuk menganalisis karakteristik masing-masing klaster yang terbentuk berdasarkan hasil clustering terhadap variabel jumlah bus dan jumlah penumpang dalam memahami perbedaan kebutuhan transportasi antar kelompok sekolah.
3. Untuk mengidentifikasi perubahan distribusi klaster dari tahun 2017-2019 dalam melihat dinamika kebutuhan transportasi sekolah dan menilai implikasi dari perubahan tersebut terhadap perencanaan dan penyediaan layanan transportasi.
4. Untuk membangun model prediktif berdasarkan hasil clustering untuk mengklasifikasikan data di tahun berikutnya.

1.3.2. Manfaat Penelitian

Berdasarkan uraian tujuan penelitian di atas, manfaat dari penelitian tersebut akan disusun ke dalam poin-poin berikut:

1. Memberikan acuan bagi pihak terkait dalam memilih algoritma clustering yang paling sesuai untuk data transportasi sekolah.

2. Memberikan pemahaman mendalam tentang pola kebutuhan transportasi di berbagai wilayah.
3. Dapat mengidentifikasi tren kebutuhan transportasi sekolah dari tahun ke tahun sebagai landasan evaluasi kebijakan transportasi sebelumnya.
4. Dapat menunjukkan potensi penggunaan model clustering sebagai alat prediktif untuk perencanaan transportasi jangka menengah dan panjang.

1.4 Metode Penelitian

Penelitian ini melibatkan beberapa metode yang digunakan untuk mendukung kelancaran dan ketepatan proses penelitian, metode tersebut akan diuraikan sebagai berikut:

1.4.1. Metode Studi Pustaka dan Literatur

Metode pustaka dan literatur merupakan kegiatan yang melibatkan sejumlah proses yaitu mengumpulkan, menelaah, dan menganalisis berbagai sumber referensi seperti buku, jurnal ilmiah, laporan penelitian, serta dokumen resmi yang relevan dengan topik yang dikaji. Sehingga, metode ini digunakan untuk memahami konsep dasar transportasi sekolah, algoritma *unsupervised learning*, serta penelitian terdahulu yang membahas analisis data transportasi. Tujuan dari metode ini adalah memperoleh landasan teori yang kuat untuk mendukung analisis data, memastikan keakuratan pendekatan yang digunakan, serta membandingkan hasil penelitian dengan temuan sebelumnya agar dapat memberikan kontribusi yang lebih komprehensif dalam bidang transportasi sekolah di Jakarta.

1.4.2. Metode Pengujian Model

Metode ini melakukan pengujian terhadap keempat model *clustering*. Pengujian dilakukan untuk mengevaluasi dari keempat model clustering untuk dipilih mana yang terbaik di antaranya. Sehingga model yang terpilih akan dianalisis untuk mengidentifikasi hubungan antara jumlah bus dan jumlah penumpang sekolah di Kota Jakarta pada periode 2017-2019. Evaluasi keempat model *clustering* akan menggunakan tiga metrik evaluasi, terdiri dari *Silhouette*

Score, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index* yang digunakan untuk menganalisis kinerja dari keempat model clustering.

1.4.3. Metode Analisis dan Kesimpulan

Hasil klaster yang terbentuk dari model *clustering* yang terpilih akan dianalisis lebih lanjut untuk mendapatkan wawasan mengenai pola hubungan jumlah bus dan jumlah penumpang sekolah di kota Jakarta dari tahun 2017-2019. Dan dari hasil tersebut akan dibentuk suatu kesimpulan terhadap hubungan yang terjadi antara jumlah bus dan jumlah penumpang sekolah di kota Jakarta dari rentang waktu yang telah ditentukan.

1.5 Sistematika Penulisan

Penulisan tugas akhir ini disusun secara berurutan dan sistematis dengan tujuan agar pembahasan menjadi terstruktur dan mudah dipahami. Berikut penulisan skripsi akan diuraikan sebagai berikut:

BAB I. PENDAHULUAN

Bagian ini terdiri dari beberapa sub-bab yaitu latar belakang, perumusan dan batasan masalah, tujuan dan manfaat penelitian, metode penelitian dan sistematika penulisan.

BAB II. TINJAUAN PUSTAKA

Bagian ini menguraikan teori yang relevan, penelitian terkait, serta konsep dasar mengenai machine learning, unsupervised learning, dan *clustering*.

BAB III. METODE PENELITIAN

Bagian ini menjelaskan sumber dataset serta alur proses *clustering*, yang mencakup pengolahan dataset, uji coba klasterisasi, dan evaluasi *clustering* dalam kaitannya dengan struktur dan kerangka kerja penelitian.

BAB IV. HASIL DAN PEMBAHASAN

Bagian ini menganalisis setiap hasil dari klaster yang terbentuk serta memvisualisasikan hasilnya.

BAB V. PENUTUP

Bagian ini menyajikan kesimpulan dari hasil *clustering* untuk menjawab pertanyaan dalam rumusan masalah serta memberikan saran bagi pengembangan penelitian selanjutnya.

DAFTAR PUSTAKA

- [1] E. Ratnawati, “MRT as an Alternative Transportation Solution which is environmentally friendly and traffic jam free in Jakarta,” in *IOP Conference Series: Earth and Environmental Science*, IOP Publishing Ltd, Jul. 2021. doi: 10.1088/1755-1315/819/1/012040.
- [2] A. Y. Nurhidayat, H. Widyastuti, Sutikno, D. P. Upahita, and A. Roschyntawati, “Impact of Traffic Volume on the Pollution Cost, Value of Time, and Travel Time Cost in Jakarta City Centre Area,” *Civil Engineering and Architecture*, vol. 11, no. 5, pp. 3209–3220, Sep. 2023, doi: 10.13189/cea.2023.110830.
- [3] G. Kreindler, A. Gaduh, T. Graff, R. Hanna, and B. A. Olken, “Optimal Public Transportation Networks: Evidence from the World’s Largest Bus Rapid Transit System in Jakarta,” 2023. doi: 10.3386/w31369.
- [4] P. Shobiroh Utami, “Analisis Model Clustering Kecelakaan Lalu Lintas Di Kota Palembang Menggunakan Pendekatan Machine Learning,” 2024.
- [5] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, “Machine Learning and Deep Learning Approaches for CyberSecurity: A Review,” 2022, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2022.3151248.
- [6] W. Jin, “Research on Machine Learning and Its Algorithms and Development,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2020. doi: 10.1088/1742-6596/1544/1/012003.
- [7] A. Lavanya *et al.*, “Assessing the Performance of Python Data Visualization Libraries: A Review,” *International Journal of Computer Engineering in Research Trends*, vol. 10, no. 1, pp. 28–39, Jan. 2023, doi: 10.22362/ijcert/2023/v10/i01/v10i0104.
- [8] V. Da Poian *et al.*, “Exploratory data analysis (EDA) machine learning approaches for ocean world analog mass spectrometry,” *Frontiers in Astronomy and Space Sciences*, vol. 10, 2023, doi: 10.3389/fspas.2023.1134141.
- [9] Z. Wang, B. He, Z. Jiang, X. Zhang, H. Dong, and D. Ye, “Distributed unsupervised meta-learning algorithm over multi-agent systems,” *Digital Communications and Networks*, Aug. 2024, doi: 10.1016/j.dcan.2024.08.006.

- [10] S. J. Ali, A. Gavric, H. Proper, and D. Bork, "Encoding Conceptual Models for Machine Learning: A Systematic Review," in *Proceedings - 2023 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion, MODELS-C 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 562–570. doi: 10.1109/MODELS-C59198.2023.00094.
- [11] M. B. Courtney, "Exploratory Data Analysis in Schools: A Logic Model to Guide Implementation," *International Journal of Education Policy and Leadership*, vol. 17, no. 4, May 2021, doi: 10.22230/ijepl.2021v17n4a1041.
- [12] M. X. Low *et al.*, "Comparison of Label Encoding and Evidence Counting for Malware Classification," *Journal of System and Management Sciences*, vol. 12, no. 6, pp. 17–30, 2022, doi: 10.33168/JSMS.2022.0602.
- [13] J. Graffelman and J. de Leeuw, "Improved Approximation and Visualization of the Correlation Matrix," *American Statistician*, vol. 77, no. 4, pp. 432–442, 2023, doi: 10.1080/00031305.2023.2186952.
- [14] Q. Yang, Q. Kang, Q. Huang, Z. Cui, Y. Bai, and H. Wei, "Linear correlation analysis of ammunition storage environment based on Pearson correlation analysis," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jun. 2021. doi: 10.1088/1742-6596/1948/1/012064.
- [15] K. Ali and A. Al-Hameed, "Spearman's correlation coefficient in statistical analysis," *Int. J. Nonlinear Anal. Appl*, vol. 13, pp. 2008–6822, 2022, doi: 10.22075/ijnaa.2022.6079.
- [16] E. F. El-Hashash and R. H. A. Shiekh, "A Comparison of the Pearson, Spearman Rank and Kendall Tau Correlation Coefficients Using Quantitative Variables," *Asian Journal of Probability and Statistics*, pp. 36–48, Oct. 2022, doi: 10.9734/ajpas/2022/v20i3425.
- [17] Z. Gu, "Complex heatmap visualization," *iMeta*, vol. 1, no. 3, Sep. 2022, doi: 10.1002/imt2.43.
- [18] Rahul Kumar Jha, "Sustainable Energy Transition: Analyzing the Impact of Renewable Energy Sources on Global Power Generation," *Journal of Artificial Intelligence and Capsule Networks*, vol. 5, no. 3, pp. 314–329, Sep. 2023, doi: 10.36548/jaicn.2023.3.007.
- [19] A. Hassan Sial, S. Yahya Shah rashdi, and Dr. A. Hafeez Khan, "Comparative Analysis of Data Visualization Libraries Matplotlib and Seaborn in Python," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 10, no. 1, pp. 277–281, Feb. 2021, doi: 10.30534/ijatcse/2021/391012021.

- [20] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, “Unsupervised Learning,” 2023, pp. 503–556. doi: 10.1007/978-3-031-38747-0_12.
- [21] M. W. B. Azlinah, M. Bee, and W. Yap, “Supervised and Unsupervised Learning for Data Science Unsupervised and Semi-Supervised Learning Series Editor: M. Emre Celebi,” 2020. doi: 10.1007/978-3-030-22475-2.
- [22] S. Naeem, A. Ali, S. Anam, and M. M. Ahmed, “An Unsupervised Machine Learning Algorithms: Comprehensive Review,” *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 911–921, 2023, doi: 10.12785/ijcds/130172.
- [23] R. Garcia-Dias, S. Vieira, W. H. Lopez Pinaya, and A. Mechelli, “Clustering analysis,” in *Machine Learning: Methods and Applications to Brain Disorders*, Elsevier, 2019, pp. 227–247. doi: 10.1016/B978-0-12-815739-8.00013-4.
- [24] K. P. Sinaga and M. S. Yang, “Unsupervised K-means clustering algorithm,” *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [25] I. Purnama Sari and I. Hanif Batubara, “Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering for Grouping Students’ Abilities in Online Learning Process,” *Journal of Computer Science, Information Technology and Telecommunication Engineering (JCoSITTE)*, vol. 2, no. 1, pp. 139–144, 2021, doi: 10.30596/jcositte.v2i1.6504.
- [26] M. Gogebakan, “A Novel Approach for Gaussian Mixture Model Clustering Based on Soft Computing Method,” *IEEE Access*, vol. 9, pp. 159987–160003, 2021, doi: 10.1109/ACCESS.2021.3130066.
- [27] Z. Kang *et al.*, “Multi-graph Fusion for Multi-view Spectral Clustering,” Sep. 2019, doi: 10.1016/j.knosys.2019.105102.
- [28] M. Shutaywi and N. N. Kachouie, “Silhouette analysis for performance evaluation in machine learning with applications to clustering,” *Entropy*, vol. 23, no. 6, Jun. 2021, doi: 10.3390/e23060759.
- [29] F. Ros, R. Riad, and S. Guillaume, “PDBI: A partitioning Davies-Bouldin index for clustering evaluation,” *Neurocomputing*, vol. 528, pp. 178–199, Apr. 2023, doi: 10.1016/j.neucom.2023.01.043.
- [30] I. Firman Ashari, E. Dwi 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,” 2023. doi: 10.30871/jaic.v7i1.4947.

- [31] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.
- [32] D. B. Hier, T. Obafemi-Ajayi, G. R. Olbricht, D. M. Burns, S. Petrenko, and D. C. Wunsch, "Enhancing Dimension-Reduced Scatter Plots with Class and Feature Centroids," Mar. 2024, [Online]. Available: <http://arxiv.org/abs/2403.20246>
- [33] A. Mazarei, R. Sousa, J. Mendes-Moreira, S. Molchanov, and H. M. Ferreira, "Online boxplot derived outlier detection," *Int J Data Sci Anal*, Jan. 2024, doi: 10.1007/s41060-024-00559-0.
- [34] Z. Koczor-Benda, T. Mateeva, and E. Rosta, "Direct Calculation of Electron Transfer Rates with the Binless Dynamic Histogram Analysis Method," *Journal of Physical Chemistry Letters*, vol. 14, no. 44, pp. 9935–9942, Nov. 2023, doi: 10.1021/acs.jpcllett.3c02624.
- [35] Sriwahyuni and R. Passarella, "Analisis Model Prediksi Untuk Layanan Bus Sekolah Jakarta Menggunakan Pendekatan Machine Learning," *Indonesian Journal of Computer Science*, 2024.