

Analisis Model Prediksi Untuk Layanan Bus Sekolah Jakarta Menggunakan Pendekatan Machine Learning

Sriwahyuni, Rossi Passarella

sw9723011@gmail.com, passarella.rossi@unsri.ac.id

Jurusan sistem Komputer, Fakultas Ilmu komputer, Universitas Sriwijaya

Informasi Artikel

Diterima : 28 Mei 2024

Direview : 5 Jun 2024

Disetujui : 30 Jun 2024

Kata Kunci

Bus sekolah, Machine learning, Prediksi, Gradient Boosting, Transportasi sekolah

Abstrak

Penelitian ini bertujuan memprediksi jenis bus sekolah di Jakarta menggunakan metode *machine learning*. Data dari tahun 2017 hingga 2019 mencakup jumlah penumpang, jumlah sekolah, dan jenis bus. Analisis data eksplorasi mengidentifikasi pola dan tren, dengan *feature engineering* menghasilkan tiga variabel utama. Tujuh model machine learning diuji, termasuk SVM, *Logistic Regression*, KNN, *Gaussian Naive Bayes*, *Decision Tree*, *AdaBoost*, dan *Gradient Boosting*, dengan fokus pada f1-score untuk menangani ketidakseimbangan data. Evaluasi menunjukkan bahwa *Gradient Boosting* memiliki performa terbaik dengan akurasi, precision, recall, dan f1-score tertinggi. Hasil penelitian memberikan wawasan tentang faktor-faktor yang mempengaruhi jenis bus sekolah dan menawarkan model prediktif yang efektif untuk mendukung pengambilan keputusan dalam pengelolaan transportasi sekolah di Jakarta. *Gradient Boosting* terbukti paling andal dalam memprediksi jenis bus sekolah, Sehingga model ini bisa menjadi dasar untuk strategi peningkatan keselamatan dan efisiensi transportasi sekolah.

Keywords

School bus, Machine learning, Prediction, Gradient Boosting, School transportation

Abstract

This research aims to predict the types of school buses in Jakarta using machine learning methods. Data from 2017 to 2019 includes the number of passengers, the number of schools, and bus types. Exploratory data analysis identified patterns and trends, with feature engineering generating three main variables. We tested seven machine learning models, including SVM, Logistic Regression, KNN, Gaussian Naive Bayes, Decision Tree, AdaBoost, and Gradient Boosting, with a focus on f1-score to handle data imbalance. The evaluation shows that gradient boosting has the best performance with the highest accuracy, precision, recall, and f1-score. The results provide insights into the factors that influence school bus types and offer an effective predictive model to support decision-making in school transportation management in Jakarta. Gradient boosting proved to be the most reliable in predicting school bus types. Therefore, this model can serve as a basis for strategies to improve the safety and efficiency of school transportation.

A. Pendahuluan

Pertumbuhan populasi yang pesat di kota-kota besar tidak hanya berarti bertambahnya jumlah manusia yang tinggal dan bekerja di kota tersebut, tetapi berimplikasi juga terhadap pertumbuhan jumlah permintaan perjalanan. Pertumbuhan permintaan perjalanan ini bukan hanya dari sisi jumlahnya saja, tetapi juga dalam panjang perjalanan karena areal kota bertambah luas seiring dengan meningkatnya kebutuhan ruang bagi kegiatan kota. Selain itu, peningkatan permintaan perjalanan ini tidak hanya berdampak pada jumlah perjalanan yang lebih tinggi, tetapi juga pada panjang perjalanan yang ditempuh. Salah satu konsekuensi dari tingginya penggunaan kendaraan pribadi sebagai sarana transportasi adalah polusi udara yang dihasilkan oleh emisi kendaraan. Istilah “polusi udara” mengacu pada keberadaan gas atau zat *aerosol* di *atmosfer* yang terakumulasi cukup banyak sehingga berdampak pada kesehatan manusia atau kesehatan ekosistem. Ini mungkin hasil dari aktivitas manusia (termasuk pertanian, industri, dan transportasi) atau peristiwa alam (termasuk kebakaran hutan). Menurut laporan dari *World Health Organization (WHO)*, dalam skala global, polusi udara semakin marak [1]. Saat ini, hanya 9% dari semua orang tinggal di daerah di mana pedoman kualitas udara WHO terpenuhi. Menurut EPA, empat sumber utama pencemaran udara adalah sumber tidak bergerak, sumber bergerak, kebakaran, dan biogenik. Transportasi, yang tercakup dalam sumber bergerak, merupakan salah satu sumber polusi udara terbesar, terutama di daerah perkotaan. Di negara-negara *Organization for Economic Co-operation and Development (OECD)*, misalnya, transportasi bertanggung jawab atas setengah dari semua emisi partikel [2].

Untuk mengatasi masalah polusi udara dan mengurangi tingkat emisi kendaraan bermotor, perlu adanya langkah-langkah pengendalian polusi yang berkelanjutan. Salah satu solusi yang dapat diimplementasikan adalah mempromosikan penggunaan kendaraan umum, seperti bus sekolah bagi para pelajar. Dengan mengoptimalkan ketersediaan dan efisiensi rute bus sekolah di Jakarta, diharapkan dapat mengurangi jumlah kendaraan pribadi yang beroperasi di jalan-jalan kota. Penggunaan kendaraan umum juga dapat membantu dalam mengurangi polusi udara karena bus sekolah biasanya beroperasi dengan kapasitas penumpang yang lebih besar dibandingkan kendaraan pribadi. Selain mengurangi tingkat polusi di udara, kendaraan pribadi juga bisa menyebabkan kemacetan di lalu lintas. Salah satu faktor utama penyebab kemacetan lalu lintas adalah volume kendaraan di jalan, yang dipengaruhi oleh beberapa faktor antara lain aksesibilitas ke berbagai tujuan dan moda transportasi. Pada penelitian [3], menyatakan bahwa terdapat lokasi-lokasi di mana orang melakukan banyak lalu lintas dan melakukan berbagai aktivitas meskipun terjadi kemacetan, yang cenderung terpusat, Kawasan terbangun dengan tingkat aksesibilitas yang lebih tinggi. Dengan meningkatkan akses terhadap transportasi umum, Masyarakat akan lebih cenderung menggunakan transportasi umum dibandingkan mengemudi sendiri, sehingga akan membantu mengurangi jumlah kendaraan di jalan dan mengurangi kemacetan.

Transportasi sekolah memainkan peran vital dalam mendukung mobilitas dan akses pendidikan bagi siswa di berbagai daerah. Di kota metropolitan seperti Jakarta, transportasi sekolah menjadi komponen penting untuk memastikan siswa tiba di sekolah dengan aman dan tepat waktu. Namun, meningkatnya jumlah siswa

dan kompleksitas jaringan transportasi perkotaan telah menimbulkan tantangan baru dalam mengelola sistem transportasi sekolah yang efektif. Minat pelajar untuk menggunakan transportasi umum sebagai moda transportasi pilihan yang digunakan untuk ke sekolah masih rendah. Hal ini disebabkan oleh beberapa faktor, termasuk *headway* yang terlalu lama antara kedatangan bus [4]. Peningkatan jumlah penumpang dan kepadatan siswa dalam bus sekolah dapat mengindikasikan tingkat keberhasilan suatu rute dan tingkat pemenuhan layanan. Di sisi lain, ketidakseimbangan antara jumlah penumpang dan kapasitas bus dapat menyebabkan ketidaknyamanan dan ketidak efisienan dalam perjalanan.

Berdasarkan penelitian [5], Beberapa temuan utama mengenai kepentingan fitur dan plot dalam hubungan *non-linier* adalah sebagai berikut: Pertama, indikator aksesibilitas menunjukkan sekitar 5 hingga 10% kepentingan fitur kecuali Mart (sekitar 50%). Kedua, aksesibilitas yang lebih baik terhadap infrastruktur transportasi umum, seperti halte bus dan stasiun transit, dikaitkan dengan lalu lintas harian rata-rata tahunan yang lebih tinggi, khususnya di wilayah metropolitan. Ketiga, akses terhadap pasar berskala besar mungkin mempunyai dampak yang tidak diinginkan terhadap volume lalu lintas kendaraan dan mobil. Keempat, terlihat bahwa tingkat lalu lintas rata-rata tahunan yang lebih rendah berhubungan dengan aksesibilitas yang lebih tinggi ke sekolah dasar untuk ketiga moda transportasi tersebut.

Ketersediaan rute bus sekolah yang memadai sangat penting untuk memastikan aksesibilitas siswa ke sekolah. Namun, di Jakarta, ketersediaan rute bus sekolah masih belum memadai [6]. Salah satu faktor yang mempengaruhi ketersediaan transportasi sekolah adalah jumlah penumpang dan kepadatan siswa. Jumlah penumpang yang terlalu banyak dan kepadatan siswa yang tinggi dapat menyebabkan keterlambatan dan ketidaknyamanan dalam perjalanan ke sekolah [7].

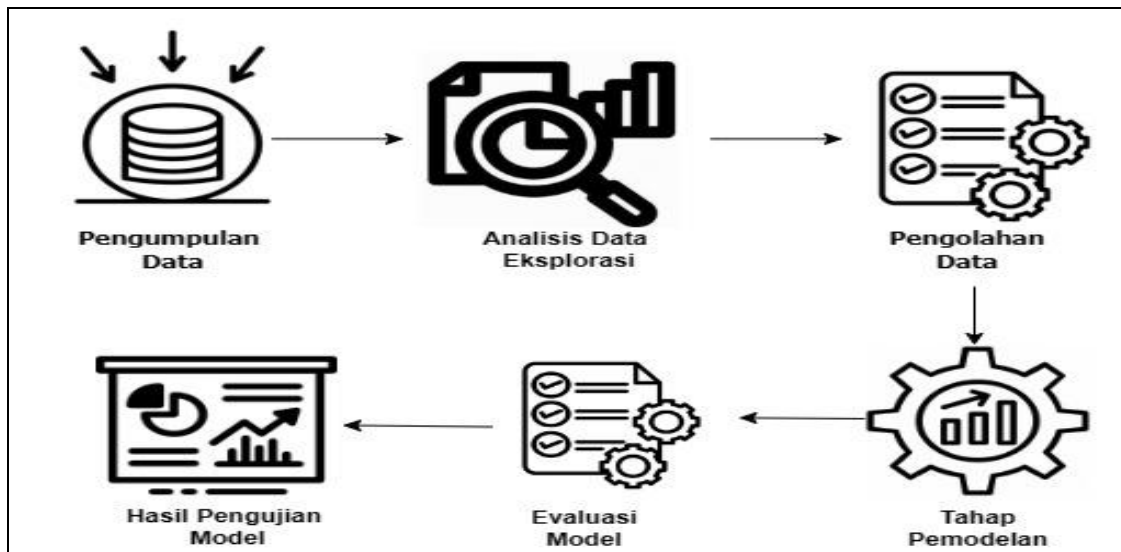
Penelitian yang dilakukan oleh [8], terkait penentuan rute dan jumlah kendaraan pada kasus *school bus routing problem*. Hasil penelitian ini memberikan analisis data tentang pengetahuan dibidang penentuan rute, baik dari pengembangan model maupun pengembangan algoritma untuk penyelesaian masalah, namun masih ada beberapa pengembangan yang harus dilakukan. seperti, penentuan nilai parameter pada travel yang lebih sesuai, atau mempertimbangkan jumlah sekolah yang dilintasi penumpang ataupun wilayah apa saja yang dilintasi pada rute bus.

Oleh sebab itu, penelitian ini dilakukan untuk mengetahui lebih dalam terkait masalah tersebut, serta mengetahui bagaimana kinerja dari model prediksi menggunakan pendekatan *machine learning* untuk memprediksi jenis bus yang optimal untuk digunakan dalam layanan bus sekolah. Sehingga, diharapkan dapat meningkatkan efisiensi, kualitas, dan kepuasan layanan bus sekolah di Jakarta melalui penggunaan jenis bus yang tepat sesuai dengan kebutuhan dan permintaan yang diprediksi.

B. Metode Penelitian

Penelitian ini menggunakan model prediksi untuk layanan bus sekolah di Jakarta, dengan menggunakan pendekatan *machine learning*. Beberapa model dikembangkan selama proses prediksi dan akan diuji pada dataset. Model-model

tersebut antara lain SVM, LR, KNN, GNB, DT, AB, dan GB. Urutan langkah-langkah penelitian diilustrasikan pada Gambar 1.



Gambar 1. Alur kerja penelitian

Tahap awal dari penelitian ini adalah pengumpulan data, yang diperoleh dari data terbuka di Jakarta [9]. Langkah selanjutnya adalah analisis data eksplorasi. Setelah itu, dilakukan prapemrosesan data, termasuk serangkaian langkah untuk membersihkan, mengelompokkan, dan menyesuaikan data mentah sebelum pemrosesan yang sebenarnya. Salah satu langkah penting dalam tahap ini adalah pembersihan data.

Data yang digunakan dalam penelitian ini diperoleh dari situs resmi Open Data Jakarta. Dataset ini berisi informasi mengenai bus sekolah di Jakarta, dengan variabel yang meliputi jenis_operasi, daerah_operasi, wilayah_operasi, jumlah_bus, dan jumlah_penumpang. Data mentah mencakup tahun 2017-2019 dan mencakup 1.009 pengamatan di lima variabel tersebut. Adapun contoh data mentah yang digunakan ditunjukkan pada gambar 2.

type_operasi	daerah_operasi	area_operasi	jumlah_bus	jumlah_penumpang
ZONASI	ZONA 1	ZONASI PONDOK GEDE - HALIM - CONDET - RANCHO	100	20358
ZONASI	ZONA 2	ZONASI KP.MELAYU - RAWAMANGUN	84	16302
ZONASI	ZONA 3	ZONASI TERMINAL KALIDERES - KAMAL	67	11416
ZONASI	ZONA 4	ZONASI KALIDERES - SEMANAN - DURIKOSAMBI	102	18655
ZONASI	ZONA 5	ZONASI PULO GADUNG - MARDANI - PASEBAN	68	12332
ZONASI	ZONA 6	ZONASI CAWANG - PEJATEN - RAGUNAN	66	6407
ZONASI	ZONA 7	ZONASI RAWAMANGUN - CIKINI - LP.BANTENG	68	10153
ZONASI	ZONA 8	ZONASI LUBANG BUAYA - DUKU.5 - KP.RAMBUTAN - RANCO	67	751
ZONASI	ZONA 9	ZONASI MARUNDA - ROROTAN	50	11794
ZONASI	ZONA 10	ZONASI MARUNDA - CIJUNING	51	6803
ZONASI	ZONA 11	ZONASI - RUSUN KAPUK - CIDENG	34	4969
REGULER	RUTE 1	LAP.BANTENG - GALLUR - P.KEMERDEKAAN	94	25549
REGULER	RUTE 2	PLUMPANG - SUNTER - KEMAYORAN	80	18334
REGULER	RUTE 3	GANDARIA - HEK - TMII	95	27055

Gambar 2. Data mentah penelitian

Selanjutnya, data mentah tersebut diproses lebih lanjut untuk membuat variabel-variabel baru, sehingga menghasilkan total 12 variabel: jenis_operasi, wilayah_operasi, daerah, kecamatan, jumlah_bus, jumlah_penumpang, tahun, bulan, km, jumlah_sekolah, sekolah_yang_dilewati, dan jenis_bus. Salah satu variabel target adalah jenis_bus. Ilustrasi dataset yang digunakan diperlihatkan pada gambar 3.

type_operasi	wilayah	kecamatan	jumlah_bus	jumlah_penumpang	tahun	bulan	km	jumlah_sekolah	sekolah_yang_dilewati	jenis_bus
ZONA 1	ZONASI PONDOK GEDE - HALIM- CONDET - R. Jakarta Timur	kramat jati - pasar ming	100	20358	2017	1	25	75	SMAN 48 – SMA/SMK U	bus kecil
ZONA 2	ZONASI KP.MELAYU - RAWAMANGUN Jakarta Timur	pulo gadung	84	16302	2017	1	27	131	SMA/SMK Cahaya Sakti	bus kecil
ZONA 3	ZONASI TERMINAL KALIDERES - KAMAL Jakarta Barat	kalideres - cangkareng	67	11416	2017	1	28	118	SMA Alamaka - SMK Syz	bus kecil
ZONA 4	ZONASI KALIDERES - SEMANAN-DURIKOSAM Jakarta Barat	kalideres - cangkareng	102	18655	2017	1	16	90	SMK Cendrawasih - SM	bus kecil
ZONA 5	ZONASI PULO GADUNG-MARDANI-PASEBAN Jakarta Timur	pulo gadung - senen	68	12332	2017	1	25	84	SMKN 17 – SMP/SMA P	bus kecil
ZONA 6	ZONASI CAWANG - PEJATEN - RAGUNAN Jakarta Selatan	kramat jati - pasar ming	69	6407	2017	1	31	52	SD/SMP/SMA Muhamr	bus kecil
ZONA 7	ZONASI RAWAMANGUN - CIKINI - LP.BANTEN Jakarta Pusat	pulo gadung - menteng	68	10153	2017	1	28	53	SD/SMP Karomiyah – Sh	bus kecil
ZONA 8	ZONASI LUBANG BUAYA - DUKU.5 - KP.RAMB Jakarta Timur	duren sawit - ciracas - k	67	751	2017	1	21	60	SD Santo Markus – SDN	bus kecil
ZONA 9	ZONASI MARUNDA - ROROTAN Jakarta Utara	cilincing	50	11794	2017	1	24	37	SDN 01,02, 04 Rorotan - bus	kecil
ZONA 10	ZONASI MARUNDA - CILINCING Jakarta Utara	cilincing	51	6803	2017	1	17	26	SMPN 244 - SMPN 266 - bus	kecil
ZONA 11	ZONASI - RUSUN KAPUK - CIDENG Jakarta Barat - Jakarta P	cengkareng - gambir	34	4969	2017	1	22	48	SMP/SMK IPPI Petojo - C	bus kecil
RUTE 1	LAP.BANTENG - GALUR -P.KEMERDEKAAN Jakarta Pusat	senen - johar baru - sav	94	25549	2017	1	24	72	SMKN 01 – SMA 01- SMI	bus sedang
RUTE 2	PLUMPANG - SUNTER - KEMAYORAN Jakarta Utara - Jakarta P	penjaringan - tanjung p	80	18334	2017	1	29	72	SMKN 54 – SMA Taman	bus sedang
RUTE 3	GANDARIA - H E K - TMII Jakarta Selatan - Jakarta	kebayoran baru - krami	95	27055	2017	1	24	127	SMK Ceger Budaya - SMI	bus sedang
RUTE 4	PERINTIS KEMERDEKAAN - PULO GADUNG- PI Jakarta Timur	cakung - pulo gadung - i	63	1943	2017	1	26	117	SMK Ristek Kikin – SMAI	bus sedang
RUTE 5	KAMPUNG MELAYU - TMII - CEGER Jakarta Timur	jatinegara - cipayang	128	36009	2017	1	26	138	SMK Ceger Budaya - SMI	bus sedang
RUTE 6	PS.MINGGU - MAMPANG PRAPATAN - KEBAY Jakarta Selatan	pasar minggu - mampai	80	18192	2017	1	25	80	SMA/SMK Bunda Kandub	bus sedang
RUTE 7	PASAR MINGGU - RANCO - LT.AGUNG - UI Jakarta Selatan - Jakarta	pasar minggu - beji	101	32202	2017	1	20	88	SMA/SMK Bunda Kandub	bus sedang

Gambar 3. Dataset yang digunakan hasil dari penambahan variabel

Adapun kategori jenis bus didasarkan pada panduan data dan informasi dari situs web data terbuka Jakarta. Secara spesifik, layanan bus sekolah dibagi menjadi dua kategori: rute reguler dan rute zonasi. Rute reguler adalah rute yang dapat dilayani oleh bus berukuran sedang dengan kapasitas penumpang 30 kursi. Rute zonasi adalah rute yang hanya dapat dilayani oleh bus kecil dengan kapasitas 19 kursi. Panduan ini disediakan dalam informasi data yang tersedia di platform data terbuka Jakarta.

C. Hasil dan Pembahasan

Bagian ini menjelaskan hasil yang diperoleh dari langkah-langkah dalam pengolahan data dan implementasi beberapa *supervised machine learning*, serta hasil analisis hasil klasifikasi.

1. Analisis Data Eksplorasi

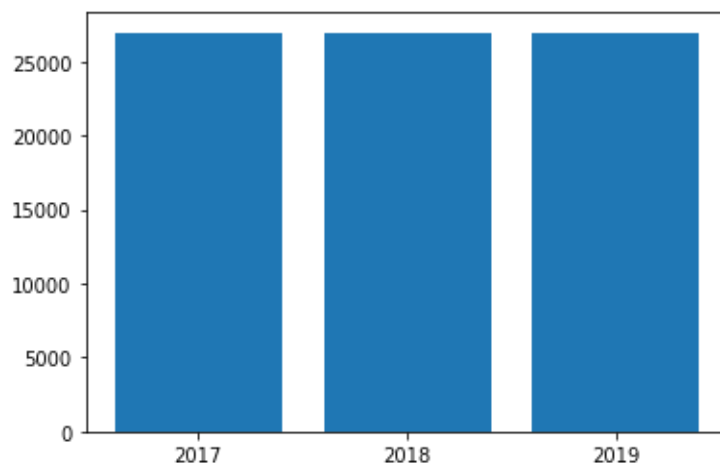
Analisis data eksploratori merupakan langkah awal dalam analisis data yang bertujuan untuk memberikan gambaran secara menyeluruh mengenai karakteristik data yang dimiliki [10]. Dalam konteks penelitian mengenai jenis bus sekolah di Jakarta, analisis data eksploratori memegang peranan penting untuk memahami berbagai aspek dan pola yang terkandung dalam data. Dalam penelitian ini, dipilih tiga variabel yang akan digunakan, yaitu jenis_bus, jumlah_sekolah, dan jumlah_penumpang. Setelah memilih 3 variabel yang digunakan untuk klasifikasi pada penelitian ini, langkah selanjutnya adalah memastikan bahwa data yang akan digunakan sudah benar dan relevan untuk penelitian. Karena tidak adanya data yang hilang (*missing value*) merupakan bagian dari proses analisis data eksploratori, berikut hasil pemeriksaan data yang dapat dilihat pada tabel 1.

Tabel 1: *Missing Value Data*

Objek	Missing Value
Jenis bus	0
Jumlah penumpang	0
Jumlah sekolah	0
Jumlah baris dalam data	1009
Jumlah kolom dalam data	3

Selanjutnya, dengan melihat pola pada diagram batang dari variabel-variabel yang digunakan dalam penelitian ini, kami akan menganalisis perbandingan pola data mengenai jumlah sekolah, jumlah penumpang, dan jenis bus sekolah dari tahun 2017 hingga 2019. Diagram batang ini dirancang untuk memberikan visualisasi yang jelas tentang perubahan selama tiga tahun ini, membantu dalam identifikasi tren dan pola yang mungkin tidak dapat dilihat melalui analisis numerik saja.

Setelah data diolah, tahap selanjutnya adalah memprediksi hasil model. Data dibagi menjadi data pelatihan dan data pengujian. Model dikembangkan dengan menggunakan data pelatihan dan kemudian diuji pada data pengujian untuk mengukur kinerjanya. Tahap selanjutnya adalah evaluasi model, yang bertujuan untuk menilai dan memilih model terbaik dari beberapa kandidat yang akan digunakan dalam penelitian ini. Setelah memilih model terbaik, langkah terakhir adalah menginterpretasikan hasil pengujian dari model yang terpilih. Berikut ini adalah pola dari variabel jumlah_sekolah pada Gambar 4.

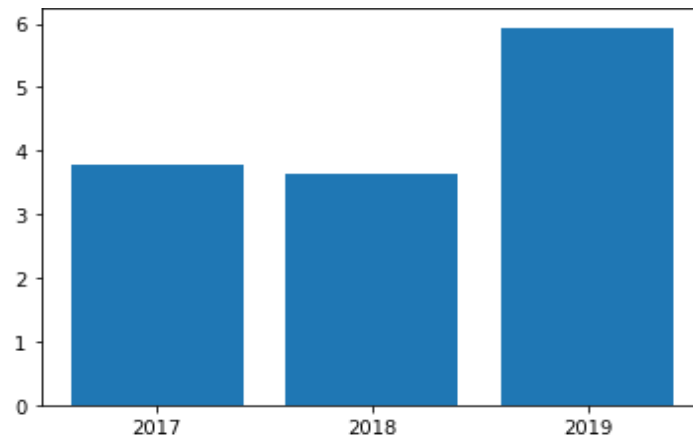
**Gambar 4.** Jumlah data sekolah pada tahun 2017-2019

Dilihat dari pola yang ada pada Gambar 4, jumlah sekolah dari tahun 2017 hingga 2019 tetap konsisten setiap tahunnya. Pola ini mengindikasikan bahwa jumlah sekolah yang mengoperasikan dan menggunakan bus sekolah tidak mengalami perubahan yang signifikan selama tiga tahun tersebut. Kestabilan jumlah sekolah ini penting untuk diperhatikan, karena memungkinkan kita untuk fokus pada variabel lain yang mungkin mempengaruhi jumlah jenis bus.

Selanjutnya, kita akan melihat diagram batang yang menggambarkan jumlah penumpang bus sekolah selama periode yang sama. Pada Gambar 5, diagram

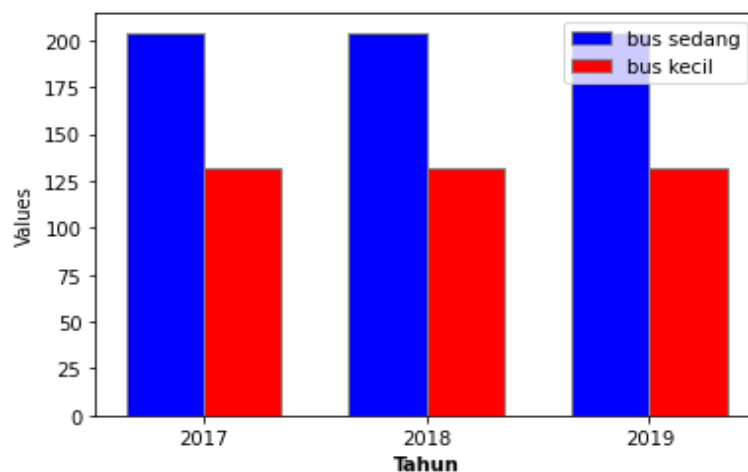
batang akan menampilkan jumlah penumpang bus sekolah untuk setiap tahun dari tahun 2017 hingga 2019. Grafik ini akan membantu kita memahami apakah ada fluktuasi jumlah penumpang yang terlibat kecelakaan dari tahun ke tahun.

Dilihat dari pola yang digambarkan pada Gambar 5, terlihat bahwa jumlah penumpang yang menggunakan bus sekolah mengalami fluktuasi setiap tahunnya. Pada tahun 2018, terjadi penurunan jumlah penumpang dibandingkan tahun 2017, namun penurunan ini tidak terlalu signifikan. Sedangkan pada tahun 2019, terjadi peningkatan jumlah penumpang yang cukup signifikan.



Gambar 5. Jumlah Penumpang tahun 2017-2019

Selanjutnya, kita akan melihat diagram yang menggambarkan jumlah jenis bus yang digunakan setiap tahunnya, seperti yang ditunjukkan pada Gambar 6. Analisis jumlah jenis bus dapat memberikan gambaran mengenai komposisi armada bus sekolah yang digunakan oleh sekolah-sekolah di Jakarta. Dalam konteks penelitian ini, diagram akan menampilkan jumlah jenis bus yang digunakan pada tahun 2017, 2018, dan 2019. Setiap batang akan mewakili satu tahun, dan tinggi batang akan mencerminkan jumlah jenis bus yang digunakan pada tahun tersebut.



Gambar 6. Jenis bus tahun 2017-2019

Dilihat dari pola yang digambarkan pada Gambar 6, terlihat bahwa jumlah jenis bus yang digunakan tidak mengalami perubahan yang signifikan dari tahun ke tahun. Konsistensi ini mengindikasikan bahwa variasi jenis bus yang digunakan oleh sekolah-sekolah di Jakarta tetap stabil selama periode pengamatan dari tahun 2017 hingga 2019.

2. Pra-pemrosesan Data

Tahap pengolahan data merupakan tahap yang krusial dan paling penting untuk mempersiapkan data sebelum dilakukan analisis lebih lanjut.

2.1. Rekayasa Fitur

Langkah awal yang dilakukan adalah memperluas dataset melalui proses rekayasa fitur. Awalnya, data mentah terdiri dari lima variabel, yang kemudian diperluas menjadi dua belas variabel. Selanjutnya, dilakukan rekayasa fitur, di mana hanya tiga variabel yang dipilih untuk model klasifikasi: *jenis_bus*, *jumlah_penumpang*, dan *jumlah_sekolah*. Ketiga variabel ini dipilih berdasarkan klasifikasi statistik mereka, seperti yang ditunjukkan pada tabel 2.

Tabel 2 : Statistik variabel dataset

Nama	Jenis data	mean	std	min	Max
jumlah_penumpang	Numerik	13263,17	9848,67	7,180	58313,00
jumlah_sekolah	numerik	80,25	34,220	26	149
Jenis_bus	Kategorikal	-	-	-	Bus sedang

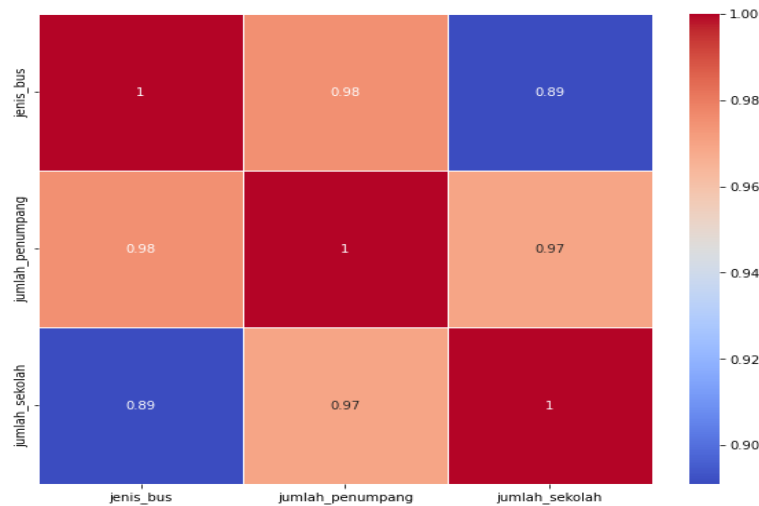
3.2.2 Pemilihan Fitur

Dalam proses prapemrosesan data, salah satu tahapannya adalah pemilihan fitur. Seleksi fitur bertujuan untuk mengidentifikasi subset fitur yang paling relevan atau penting untuk digunakan dalam membangun model pembelajaran mesin. Dengan memilih fitur terbaik di awal, dimungkinkan untuk mengurangi dimensi data dan meningkatkan efisiensi dan kinerja model. Hal ini dikarenakan model hanya akan mempertimbangkan fitur-fitur yang memiliki dampak signifikan dalam memprediksi variabel target, dimana variabel target dalam penelitian ini adalah *type_bus*. Pada tahap ini, digunakan metode *f_classif* yang menghasilkan nilai untuk setiap fitur. Dua fitur terbaik dipilih berdasarkan nilai tersebut, seperti yang ditunjukkan pada tabel 3.

Tabel 3. Variabel *predictor*

Fitur	Skor
Jumlah_penumpang	91.770092
Jumlah_sekolah	57.491207

Berikut ini adalah bentuk matriks korelasi dari ketiga variabel yang digunakan, yaitu pada Gambar 7.



Gambar 7. Matriks Korelasi

3. Tahap Pemodelan

Pada tahap pemodelan ini, tujuh model akan diuji pada dataset: SVM, LR, KNN, GNB, DT, AB, dan GB. Untuk variabel target, digunakan variabel bus_type, dengan kelas “0” yang mewakili bus kecil dan kelas “1” yang mewakili bus sedang. Karena variabel bus_type hanya memiliki dua kategori, variabel ini disebut sebagai kelas biner, dengan dua kondisi: benar dan salah. Distribusi variabel target dalam Python untuk variabel bus_type terdiri dari total 1.008 titik data. Dapat diamati bahwa kelas-kelas tersebut tidak seimbang (*imbalanced data*), di mana kelas “0” terdiri dari 396 titik data, sedangkan kelas “1” terdiri dari 612 titik data, seperti yang ditunjukkan pada tabel 4. Untuk alasan ini, fokus pemilihan model adalah nilai *f1-score*.

Tabel 4. Distribusi dataset variabel target

Total Data	Jenis_bus	
	0	1
1.008	396	612

4. Evaluasi Model

Pada tahap evaluasi ini, akan dilakukan pemilihan di antara beberapa model yang telah diuji sebelumnya untuk menentukan model terbaik yang akan digunakan dalam penelitian. Proses ini sangat penting untuk memastikan bahwa model yang dipilih mampu memberikan prediksi yang akurat dan sesuai dengan tujuan penelitian [11].

4.1 Model Menggunakan Data *Training*

Perbandingan model dari ketujuh metode *machine learning* untuk data training dapat dilihat pada tabel 5.

Tabel 4. Perbandingan matrix evaluasi data training

Model	Accuracy	Jenis Bus	Precision	Recall	F1-score
SVM	0,68	0	0,60	0,54	0,57
		1	0,72	0,77	0,74
		Mean	0,66	0,65	0,65
LR	0,69	0	0,64	0,49	0,56
		1	0,71	0,81	0,76
		Mean	0,67	0,65	0,66
KNN	0,90	0	0,87	0,88	0,87
		1	0,92	0,91	0,92
		Mean	0,89	0,89	0,89
GNB	0,64	0	0,54	0,63	0,53
		1	0,73	0,64	0,68
		Mean	0,63	0,64	0,63
DT	0,80	0	0,97	0,52	0,68
		1	0,76	0,99	0,86
		Mean	0,86	0,75	0,77
AB	0,96	0	0,94	0,97	0,95
		1	0,98	0,96	0,97
		Mean	0,96	0,96	0,96
GB	0,98	0	0,95	0,99	0,97
		1	0,99	0,97	0,98
		Mean	0,97	0,98	0,97

Analisis masing-masing model berdasarkan metrik evaluasi pada data training di atas, yaitu pertama, model SVM memiliki akurasi sebesar 0.68 pada data training. *Precision*, *recall*, dan *F1-score* untuk kelas 0 (bus kecil) adalah 0.60, 0.54, dan 0.57, sedangkan untuk kelas 1 (bus sedang) adalah 0.72, 0.77, dan 0.74. sedangkan rata-rata *precision*, *recall*, dan *F1-score* adalah 0.66. dapat disimpulkan bahwa SVM memiliki performa yang cukup rendah dibanding dengan model yang lain.

Kemudian model LR mendapatkan hasil analisis yaitu LR memiliki akurasi sebesar 0.69 pada data training. Nilai *precision*, *recall*, dan *F1-score* untuk kelas 0 adalah 0.64, 0.49, dan 0.56, sedangkan untuk kelas 1 adalah 0.71, 0.81, dan 0.76. sedangkan nilai rata-rata *precision*, *recall*, dan *F1-score* adalah 0.67. maka dapat disimpulkan bahwa LR menunjukkan performa yang sedikit lebih baik dibandingkan dengan SVM namun masih kurang baik.

Selanjutnya, model KNN memiliki akurasi sebesar 0.90 pada data training. Nilai *precision*, *recall* dan *F1-score* untuk kelas 0 adalah 0.87, 0.88 dan 0.87, sedangkan untuk kelas 1 adalah 0.92, 0.91 dan 0.92. Kemudian rata-rata *precision*, *recall*, dan *F1-score* adalah 0.89.

Kemudian, model GNB memiliki akurasi sebesar 0.64 pada data training. Selain itu, nilai *precision*, *recall*, dan *F1-score* untuk kelas 0 adalah 0.54, 0.63, dan 0.53, sedangkan untuk kelas 1 adalah 0.73, 0.64, dan 0.68. Kemudian rata-rata *precision*, *recall*, dan *F1-score* adalah 0.63. Sehingga dapat disimpulkan bahwa GNB memiliki performa yang cukup rendah, dan merupakan model dengan performa terburuk diantara yang lain.

Model DT memiliki akurasi sebesar 0.80 pada data training. Nilai *precision*, *recall* dan *F1-score* untuk kelas 0 adalah 0.97, 0.52 dan 0.68, sedangkan untuk kelas 1 adalah 0.76, 0.99 dan 0.86. Kemudian nilai rata-rata *precision*, *recall* dan *F1-score* adalah 0.79.

Model AB memiliki akurasi sebesar 0.96 pada data training. Nilai *precision*, *recall*, dan *F1-score* untuk kelas "0" adalah 0.94, 0.97, dan 0.95, sedangkan untuk kelas "1" adalah 0.98, 0.96, dan 0.97. maka nilai rata-rata *F1-score* adalah 0.96. dapat disimpulkan bahwa AB memiliki performa yang baik.

Model GB memiliki akurasi sebesar 0.98 pada data training. Nilai *precision*, *recall*, dan *F1-score* untuk kelas "0" adalah 0.95, 0.99, dan 0.97, sedangkan untuk kelas "1" adalah 0.99, 0.97, dan 0.98. maka rata-rata nilai *F1-score* adalah 0.97. maka dapat disimpulkan bahwa GB memiliki performa yang sangat baik.

Dari analisis diatas dapat disimpulkan bahwa model AB merupakan model terbaik untuk mengklasifikasikan jenis bus pada dataset ini. AB memberikan performa yang baik dengan akurasi yang hampir sempurna dan nilai *precision*, *recall*, dan *F1-score* yang tinggi. Hal ini diikuti oleh GB yang juga menunjukkan performa yang sangat baik. Pada tabel 6, ditunjukkan hasil dari 2 model terbaik dari data *training*.

Tabel 5. Performa f1-score 2 model terbaik data *training*

Model	Bus kecil (0)	Bus sedang (1)
AB	0,95	0,97
GB	0,97	0,98

4.2 Model Menggunakan Data *Testing*

Perbandingan model dari ketujuh metode machine learning untuk data testing dapat dilihat pada tabel 7.

Analisis dari masing-masing model diperoleh berdasarkan metrik evaluasi dari ketujuh model yang digunakan, yaitu model SVM memiliki akurasi sebesar 0.68 pada data testing. *Precision* dan *recall* untuk kelas 0 (bus kecil) adalah 0.57 dan 0.59, sedangkan untuk kelas 1 (bus sedang) adalah 0.75 dan 0.74. Nilai *F1-score* untuk kelas 0 adalah 0.58 dan untuk kelas 1 adalah 0.75. Model SVM menunjukkan performa yang rendah dibandingkan dengan model lainnya. Dapat disimpulkan bahwa SVM tidak cukup baik dalam memprediksi kelas bus kecil dan bus sedang pada dataset ini. Selanjutnya, model LR memiliki akurasi sebesar 0.68 pada data *testing*. *Precision* dan *recall* untuk kelas 0 adalah 0.58 dan 0.48, sedangkan untuk kelas 1 adalah 0.72 dan 0.80. *F1-score* untuk kelas 0 adalah 0.53 dan untuk kelas 1 adalah 0.76. Performa LR sedikit lebih baik daripada SVM tetapi masih rendah. Dapat disimpulkan bahwa LR juga tidak cukup baik dalam memprediksi kelas bus kecil dan bus sedang pada dataset ini.

Adapun, model KNN memiliki akurasi sebesar 0.77 pada data testing. *Precision* dan *recall* untuk kelas 0 adalah 0.75 dan 0.58, sedangkan untuk kelas 1 adalah 0.77 dan 0.89. Nilai *F1-score* untuk kelas 0 adalah 0.64 dan untuk kelas 1 adalah 0.83. KNN menunjukkan performa yang lebih baik daripada SVM dan LR.

Dapat disimpulkan bahwa KNN memiliki performa yang cukup baik dalam memprediksi kelas bus kecil dan bus sedang.

Tabel 6. Perbandingan *matrix* evaluasi data *testing*

Model	Accuracy	Jenis Bus	Precision	Recall	F1-score
SVM	0,68	0	0,57	0,59	0,58
		1	0,75	0,74	0,75
		Mean	0,66	0,66	0,66
LR	0,68	0	0,58	0,48	0,53
		1	0,72	0,80	0,76
		Mean	0,65	0,64	0,64
KNN	0,77	0	0,75	0,58	0,64
		1	0,77	0,89	0,83
		Mean	0,76	0,72	0,73
GNB	0,66	0	0,54	0,65	0,59
		1	0,77	0,67	0,71
		Mean	0,65	0,66	0,65
DT	0,93	0	0,93	0,88	0,90
		1	0,93	0,96	0,95
		Mean	0,93	0,92	0,92
AB	0,95	0	0,91	0,96	0,94
		1	0,98	0,94	0,96
		Mean	0,94	0,95	0,95
GB	0,98	0	0,92	0,97	0,95
		1	0,98	0,95	0,97
		Mean	0,95	0,96	0,96

Sedangkan, model GNB memiliki akurasi 0,66 pada data testing. Precision dan recall untuk kelas 0 adalah 0.54 dan 0.65, sedangkan untuk kelas 1 adalah 0.77 dan 0.67. Nilai *F1-score* untuk kelas 0 adalah 0.59 dan untuk kelas 1 adalah 0.71. GNB memiliki kinerja yang relatif rendah dibandingkan dengan model lainnya. Dapat disimpulkan bahwa GNB juga tidak cukup baik dalam memprediksi kelas bus kecil dan bus sedang pada dataset ini. Model DT memiliki akurasi sebesar 0.93 pada data testing. Precision dan recall untuk kelas 0 adalah 0.93 dan 0.88, sedangkan untuk kelas 1 adalah 0.93 dan 0.96. Nilai *F1-score* untuk kelas 0 adalah 0.90 dan untuk kelas 1 adalah 0.95. Model AB memiliki akurasi sebesar 0.95 pada data testing. Presesi memiliki nilai rata-rata 0.94. Nilai recall memiliki nilai rata-rata sebesar 0.95. Sedangkan nilai *f1-Score* memiliki nilai mean sebesar 0.95.

Model GB memiliki nilai akurasi sebesar 0.96 pada data testing. Presesi memiliki nilai mean sebesar 0.95. Nilai *recall* memiliki nilai mean sebesar 0.96. Sedangkan nilai *f1-Score* memiliki nilai rata-rata sebesar 0.96. Dengan memperhatikan performa yang diberikan oleh masing-masing model pada data testing, dapat disimpulkan bahwa AB dan GB merupakan model yang terbaik dalam memprediksi jenis bus (bus kecil dan bus sedang) pada dataset ini. Model ini memberikan performa yang sangat baik dan menunjukkan nilai *fitting* yang baik. Tabel 8 menunjukkan hasil dari 2 model terbaik dari data *testing*.

Tabel 7. Performa f1-score 2 model terbaik data *testing*

Model	Bus kecil (0)	Bus sedang (1)
AB	0,94	0,96
GB	0,95	0,97

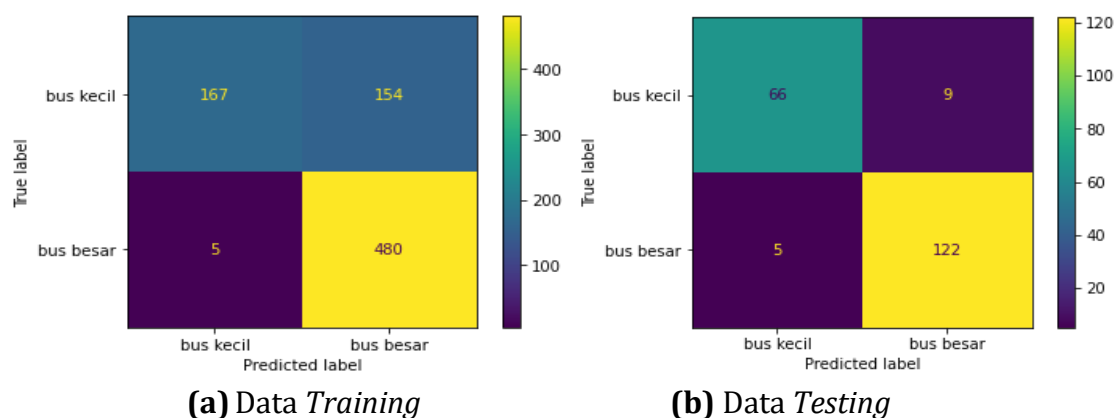
5. Hasil Pengujian

Didapatkan hasil model terbaik yaitu model AB dan GB. Untuk mengetahui sejauh mana performa yang dihasilkan oleh 2 model machine learning terbaik dalam mengklasifikasikan jenis bus dengan menggunakan data training dan data testing, maka akan dilakukan analisa lebih lanjut, untuk menentukan model terbaik berdasarkan perbandingan nilai rata-rata yang dihasilkan oleh masing-masing model. Perbandingan hasil data training dan data testing seperti yang ditunjukkan pada tabel 9.

Dari hasil performa model pada data training dan data testing, ada beberapa hal yang dapat diamati, yaitu pada AB. Pada data training, AB memiliki akurasi sebesar 96%, dengan nilai presisi sebesar 94% untuk kelas 0 (bus kecil) dan 98% untuk kelas 1 (bus yang ada saat ini). Pada data testing, performa AB mencapai akurasi 95%, memiliki nilai presisi untuk kelas “1” sebesar (98%), namun presisi untuk kelas “0” sebesar 91%.

Tabel 8. Hasil performa model data training dan data testing

Model	Data Training			Data Testing		
	0	1	Mean	0	1	Mean
AB	0,95	0,97	0,96	0,94	0,96	0,95
GB	0,97	0,98	0,97	0,95	0,97	0,96

**Gambar 8.** Hasil klasifikasi data *training* dan data *testing* model GB

Sedangkan pada model GB untuk data *training*, GB memiliki akurasi yang lebih tinggi dibandingkan AB yaitu 98%, presisi untuk kelas “1” (99%) lebih tinggi dibandingkan AB. Pada data testing, performa GB memiliki nilai dengan akurasi 96%. Presisi untuk kedua kelas mencapai 92% untuk kelas “0” dan 98% untuk kelas “1”. Untuk nilai f1-score, GB mendapatkan nilai 96% untuk nilai testing. Sedangkan model AB mendapatkan nilai yang lebih kecil untuk nilai f1-score pada data testing, yaitu 95%.

Oleh karena itu, dapat disimpulkan bahwa model AB memiliki performa yang lebih rendah pada data training dan *testing*. Di sisi lain, GB memiliki performa yang lebih baik daripada model AB. GB menunjukkan kemampuan generalisasi yang lebih baik. Berdasarkan hasil tersebut, GB direkomendasikan untuk digunakan pada penelitian ini karena mampu memberikan performa yang lebih stabil antara data training dan testing. Gambar 8 merupakan hasil klasifikasi data training dan data testing. Dari 7 model yang diuji, 1 model terbaik yang didapatkan yaitu model GB merupakan model terbaik yang telah didapatkan.

D. Kesimpulan

Dalam penelitian ini, tujuan utamanya adalah untuk memprediksi jenis bus yang optimal untuk digunakan dalam layanan bus sekolah. Dengan menggunakan berbagai model machine learning, menentukan model terbaik dalam memprediksi jenis bus yang tepat berdasarkan fitur-fitur yang ada. Berikut ini adalah penjelasan sesuai dengan hasil penelitian yang telah dilakukan:

- Pilihan Model Terbaik Dari hasil evaluasi, GB menunjukkan performa terbaik dalam memprediksi jenis bus, dengan akurasi 96% pada data testing. GB memiliki kemampuan untuk menghasilkan aturan keputusan yang mudah dipahami, sehingga cocok untuk aplikasi praktis seperti prediksi jenis bus.
- Analisis Performa Model terbaik yaitu model AB memiliki performa yang baik pada data *training* dan data *testing*. Namun, nilai model AB masih berada di bawah nilai model GB. Model AB cukup baik karena memiliki hasil *fitting* yang baik. Sedangkan model (GB) memiliki performa *fitting* yang baik, baik dari data training maupun data *testing*. Model GB memiliki hasil yang paling baik dari keenam model lainnya. GB juga memberikan keuntungan dalam hal interpretasi model, karena aturan keputusan yang dihasilkan dapat dengan mudah dimengerti oleh pengguna.

Berdasarkan penelitian ini, dapat disimpulkan bahwa model GB merupakan model yang paling sesuai untuk memprediksi jenis bus yang optimal untuk digunakan dalam pelayanan bus sekolah. Model ini dapat memberikan prediksi dengan akurasi yang tinggi dan mudah dipahami, sehingga cocok untuk diimplementasikan dalam keputusan nyata terkait layanan bus sekolah. Dengan demikian, untuk jenis bus yang digunakan, jenis bus yang lebih baik dilihat dari penelitian yang telah dijelaskan di atas. Dan pemilihan model tergantung pada kebutuhan spesifik dari kasus penggunaan, dengan pertimbangan yang cermat terhadap kinerja model, kompleksitas, dan interpretabilitas.

E. Referensi

- [1] P. Sicard, E. Agathokleous, A. De Marco, E. Paoletti, and V. Calatayud, "Urban population exposure to air pollution in Europe over the last decades," *Environ. Sci. Eur.*, vol. 33, no. 1, p. 28, 2021, doi: 10.1186/s12302-020-00450-2.
- [2] D. L. Mendoza *et al.*, "Idle-Free Campaign Survey Results and Idling Reductions in an Elementary School," *Vehicles*, vol. 4, no. 3, pp. 865–902, 2022, doi: 10.3390/vehicles4030048.
- [3] A. Mondschein and B. Taylor, "Is traffic congestion overrated? Examining the highly variable effects of congestion on travel and accessibility," *J. Transp. Geogr.*, vol. 64, pp. 65–76, Oct. 2017, doi: 10.1016/j.jtrangeo.2017.08.007.
- [4] B. B. Perdana, "Perencanaan Angkutan Sekolah Di Kabupaten Bantul," 2022, [Online]. Available: [http://digilib.ptdisttd.net/2416/%0Ahttp://digilib.ptdisttd.net/2416/1/Draft KKW Final %28Bhima Perdana%29.pdf](http://digilib.ptdisttd.net/2416/%0Ahttp://digilib.ptdisttd.net/2416/1/Draft%20KKW%20Final%20Bhima%20Perdana.pdf)
- [5] S. Lee, J. Yang, and K. Cho, "The Influence of Transportation Accessibility on Traffic Volumes in South Korea : An Extreme Gradient Boosting Approach," 2023.
- [6] F. L. Nabilah, S. Anggada, S. K. Koto, and R. P. Naskah, "Perencanaan angkutan sekolah di kabupaten landak," no. 5, pp. 1–15, 2022.
- [7] N. A. K. Sakti and S. T. Pamungkas, "Kondisi Fisik Jalur Pejalan Kaki pada Koridor Jalan Veteran Malang dalam Konteks Kenyamanan Spasial," *J. Mhs. Jur. Arsit. Univ. Brawijaya*, vol. 7, no. 3, 2019.
- [8] F. Ramadhan and A. Imran, "Penentuan Rute Dan Jumlah Kendaraan Pada Kasus School Bus Routing Problem: Penerapan Algoritma Record-to-record Travel," *J. Rekayasa Sist. Ind.*, vol. 5, no. 01, p. 15, 2018, doi: 10.25124/jrsi.v5i01.307.
- [9] R. Reguler and B. U. S. Sedang, "UNIT PENGELOLA ANGKUTAN SEKOLAH DINAS PERHUBUNGAN PROVINSI DKI JAKARTA Rute Lintasan Reguler," 2017.
- [10] V. Yellapu, "Full Text Introduction," *Descr. Stat.*, vol. 4, no. 1, pp. 60–63, 2018, doi: 10.4103/IJAM.IJAM.
- [11] S. Peng, J. Zhu, Z. Liu, B. Hu, M. Wang, and S. Pu, "Prediction of Ammonia Concentration in a Pig House Based on Machine Learning Models and Environmental Parameters," *Animals*, vol. 13, no. 1, pp. 1–14, 2023, doi: 10.3390/ani13010165.