

Analisis Data Jumlah Pemohon Perpanjangan SIM Dengan Metode Crisp-DM Sebagai Rekomendasi Penjadwalan Layanan SIM Keliling (Studi Kasus: Polres Metro Depok)

Athaya Rima Hariyanto^{1*}, Leni Susanti²

¹²Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, Banten, Indonesia

*Corresponding Author Email: athaya.rima@gmail.com

Abstract

The Mobile SIM extension service at Polres Metro Depok faces challenges related to long queues and static operational scheduling that does not fully accommodate fluctuations in applicant demand. This study aims to analyze the pattern of SIM extension applicants as a basis for recommending a more flexible and effective Mobile SIM service scheduling strategy. The study applies a descriptive quantitative approach using the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. The dataset consists of secondary historical data covering 11,756 SIM A and SIM C applicants recorded during January–June 2025 across 131 operational days. Data processing includes data understanding, data preparation, K-Means clustering, model evaluation, and deployment of the analysis results into a web-based dashboard prototype. K-Means was applied with three clusters representing low, medium, and high service density. The cluster quality was evaluated using the Davies-Bouldin Index (DBI) through RapidMiner and Python. RapidMiner produced a DBI value of 0.583, while Python produced 0.520. The results indicate substantial fluctuations in applicant demand, with the lowest monthly average occurring in April at 36.76 applicants and the highest in June at 166.68 applicants. The clustering results were subsequently transformed into an adaptive operational scheduling recommendation and implemented in a Python-Flask web dashboard prototype.

Keywords: Mobile SIM, CRISP-DM, K-Means Clustering, Dynamic Scheduling, Polres Metro Depok

Abstrak

Pelayanan perpanjangan SIM Keliling di Polres Metro Depok menghadapi kendala berupa antrean panjang dan jadwal operasional yang masih bersifat statis sehingga belum sepenuhnya menyesuaikan fluktuasi jumlah pemohon. Penelitian ini bertujuan menganalisis pola jumlah pemohon perpanjangan SIM sebagai dasar penyusunan rekomendasi penjadwalan layanan SIM Keliling yang lebih fleksibel dan efektif. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan kerangka kerja Cross-Industry Standard Process for Data Mining (CRISP-DM). Data yang digunakan merupakan data sekunder historis sebanyak 11.756 pemohon SIM A dan SIM C selama Januari–Juni 2025 pada 131 hari operasional. Tahapan penelitian meliputi pemahaman data, persiapan data, pemodelan menggunakan algoritma K-Means, evaluasi, dan deployment hasil analisis ke dalam prototipe dasbor berbasis web. K-Means diterapkan dengan tiga klaster yang merepresentasikan tingkat kepadatan sepi, sedang, dan ramai. Evaluasi kualitas klaster menggunakan Davies-Bouldin Index (DBI) menghasilkan nilai 0,583 pada RapidMiner dan 0,520 pada Python. Hasil analisis menunjukkan adanya fluktuasi permintaan yang cukup tinggi, dengan rata-rata terendah pada April sebesar 36,76 pemohon dan tertinggi pada Juni sebesar 166,68 pemohon. Hasil klasterisasi kemudian digunakan sebagai dasar rekomendasi penjadwalan operasional adaptif dan diimplementasikan dalam prototipe dasbor berbasis Python dan Flask.

Kata Kunci: SIM Keliling, CRISP-DM, K-Means Clustering, Penjadwalan Dinamis, Polres Metro Depok

A. PENDAHULUAN

Pelayanan publik merupakan salah satu bentuk tanggung jawab pemerintah kepada masyarakat, termasuk pelayanan penerbitan dan perpanjangan Surat Izin Mengemudi (SIM) oleh Kepolisian Negara Republik Indonesia. Salah satu upaya untuk mempermudah masyarakat adalah layanan SIM Keliling. Efektivitas layanan ini dipengaruhi oleh

kesesuaian jadwal, lokasi, kapasitas pelayanan, dan jumlah pemohon.

Pada wilayah Polres Metro Depok, jumlah pemohon perpanjangan SIM mengalami perubahan selama Januari–Juni 2025, yaitu 2.105 pemohon pada Januari, 2.080 pada Februari, 1.261 pada Maret, 619 pada April, 2.024 pada Mei, dan meningkat menjadi 3.667 pada Juni. Perubahan tersebut menunjukkan bahwa kebutuhan layanan tidak

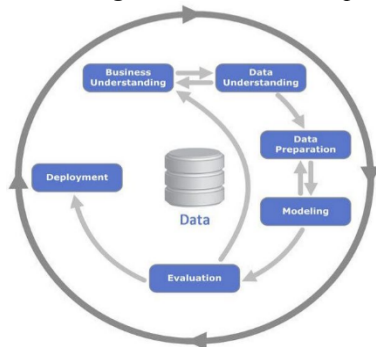
bersifat konstan sehingga penjadwalan yang statis berpotensi menimbulkan ketidaksesuaian antara kapasitas pelayanan dan jumlah pemohon.

Data historis tersebut dapat dianalisis menggunakan data mining untuk memperoleh pola kepadatan pelayanan. Penelitian ini menggunakan kerangka CRISP-DM dan algoritma K-Means untuk mengelompokkan tingkat kepadatan pemohon menjadi kategori sepi, sedang, dan ramai. Penelitian terdahulu telah menerapkan CRISP-DM dan K-Means untuk berbagai kebutuhan clustering, sedangkan penelitian mengenai SIM Keliling lebih banyak membahas aspek kualitas pelayanan. Perbedaan penelitian ini terletak pada pemanfaatan hasil clustering sebagai dasar rekomendasi penjadwalan layanan SIM Keliling.

Penelitian ini bertujuan menganalisis jumlah pemohon perpanjangan SIM, mengidentifikasi tingkat kepadatan pelayanan menggunakan K-Means, mengevaluasi hasil clustering, serta menghasilkan rekomendasi penjadwalan SIM Keliling yang lebih adaptif. Hasil analisis selanjutnya diwujudkan dalam prototipe dasbor berbasis web sebagai bentuk deployment.

B. METODE

Penelitian ini merupakan penelitian terapan dengan pendekatan kuantitatif. Penelitian menggunakan data sekunder berupa data jumlah pemohon perpanjangan SIM A dan SIM C pada layanan SIM Keliling Polres Metro Depok periode Januari sampai Juni 2025. Data yang digunakan berjumlah 11.756 pemohon yang tersebar pada 131 hari operasional. Metode analisis yang digunakan adalah **Cross-Industry Standard Process for Data Mining (CRISP-DM)**, yang terdiri dari enam tahapan, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*.



Gambar 1. Tahapan CRISP DM

Business Understanding

Tahap *Business Understanding* dilakukan untuk mengidentifikasi permasalahan pada pelayanan SIM Keliling Polres Metro Depok. Permasalahan utama yang ditemukan adalah jumlah pemohon yang mengalami fluktuasi sehingga beban pelayanan pada setiap hari dan lokasi tidak selalu sama. Sementara itu, jadwal operasional yang digunakan masih bersifat tetap dan belum sepenuhnya mempertimbangkan pola jumlah pemohon. Oleh karena itu, penelitian diarahkan untuk menghasilkan pengelompokan tingkat kepadatan pemohon sebagai dasar rekomendasi penjadwalan layanan SIM Keliling.

Data Understanding

Tahap *Data Understanding* dilakukan untuk memahami karakteristik data yang diperoleh. Data yang digunakan merupakan data historis pemohon perpanjangan SIM selama Januari sampai Juni 2025. Pada tahap ini dilakukan pemeriksaan terhadap struktur data, jumlah pemohon, tanggal pelayanan, serta lokasi pelayanan. Analisis awal juga dilakukan untuk mengetahui perubahan jumlah pemohon selama periode penelitian dan mengidentifikasi kondisi data yang perlu dipersiapkan sebelum proses pemodelan.

Data Preparation

Tahap *Data Preparation* dilakukan untuk menyiapkan data agar dapat digunakan dalam proses analisis. Proses yang dilakukan meliputi seleksi atribut, pembersihan data, konstruksi data, dan penyesuaian format data. Atribut yang digunakan dalam penelitian meliputi tanggal, jumlah pemohon, dan lokasi pelayanan. Data dengan nilai nol yang menunjukkan hari tidak beroperasi atau hari libur dikeluarkan sehingga data yang digunakan dalam proses analisis merupakan data pada hari pelayanan aktif.

Modeling

Tahap *Modeling* dilakukan dengan menerapkan algoritma **K-Means Clustering** untuk mengelompokkan tingkat kepadatan jumlah pemohon. Jumlah kluster yang digunakan adalah tiga kluster ($K=3$), yaitu **Sepi, Sedang, dan Ramai**. Proses pengelompokan dilakukan berdasarkan karakteristik jumlah pemohon pada setiap hari pelayanan. Pemodelan dilakukan secara manual dan kemudian dibandingkan menggunakan perangkat RapidMiner dan Python untuk melihat konsistensi hasil pengelompokan.

Evaluation

Tahap *Evaluation* dilakukan untuk mengetahui kualitas hasil pengelompokan yang telah diperoleh. Evaluasi menggunakan **Davies-Bouldin Index (DBI)** sebagai ukuran kualitas kluster. Selain itu, hasil pengelompokan dari perhitungan manual dibandingkan dengan hasil yang diperoleh melalui RapidMiner dan Python. Tahap ini bertujuan memastikan bahwa hasil clustering dapat digunakan sebagai dasar dalam penyusunan rekomendasi penjadwalan layanan SIM Keliling.

Deployment

Tahap *Deployment* merupakan tahap penerapan hasil analisis ke dalam bentuk rekomendasi. Hasil pengelompokan tingkat kepadatan pemohon digunakan sebagai dasar dalam menentukan prioritas operasional layanan SIM Keliling. Kluster Ramai menjadi perhatian utama dalam pengaturan pelayanan karena menunjukkan tingkat kepadatan pemohon yang lebih tinggi, sedangkan kluster Sedang dan Sepi digunakan sebagai dasar untuk menyesuaikan pengelolaan operasional sesuai dengan kondisi jumlah pemohon.

C. HASIL DAN PEMBAHASAN

Hasil Data Understanding

Hasil pengumpulan data menunjukkan bahwa jumlah pemohon mengalami perubahan pada setiap bulan. Jumlah pemohon pada Januari sebanyak 2.105 orang, Februari

2.080 orang, Maret 1.261 orang, April 619 orang, Mei 2.024 orang, dan Juni 3.667 orang. Total keseluruhan mencapai 11.756 pemohon.

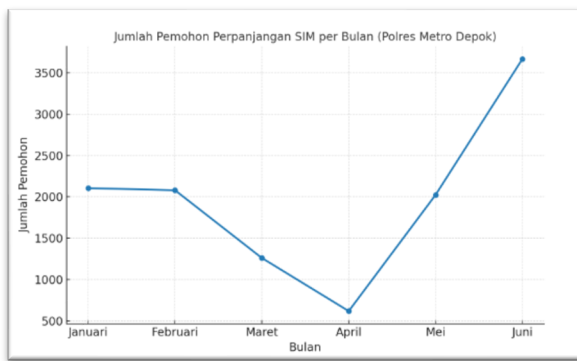
Analisis statistik harian menunjukkan adanya perbedaan rata-rata jumlah pemohon antarbulan. Rata-rata tertinggi terdapat pada Juni sebesar 166,68 pemohon per hari, sedangkan rata-rata terendah terdapat pada April sebesar 36,76 pemohon per hari. Nilai minimum dan maksimum juga menunjukkan variasi volume pelayanan, sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Statistik Deskriptif Jumlah Pemohon

Bulan	Min (>0)	Max	Rata-rata (>0)
Januari	29	215	91,52
Februari	31	198	86,66
Maret	28	176	54,82
April	30	165	36,76
Mei	35	202	85,65
Juni	40	245	166,68

Sumber: Data penelitian, 2025.

Variasi tersebut menunjukkan bahwa jumlah pemohon tidak bersifat konstan. Perbedaan volume pelayanan menjadi dasar dilakukannya proses clustering untuk memperoleh kelompok tingkat kepadatan.



Gambar 2. Grafik Jumlah Pemohon

Gambar tersebut memperlihatkan bahwa rata-rata jumlah pemohon mengalami penurunan hingga April dan meningkat kembali pada Mei, kemudian mencapai nilai tertinggi pada Juni. Pola tersebut memperkuat kebutuhan untuk melakukan pengelompokan berdasarkan tingkat kepadatan pelayanan.

Hasil Data Preparation

Data yang telah dipilih terdiri atas tanggal, jumlah pemohon, dan lokasi pelayanan. Pada tahap *clean data*, nilai nol yang merepresentasikan hari tidak beroperasi tidak dimasukkan dalam proses clustering. Selanjutnya data disusun kembali agar hanya memuat data operasional aktif dan berada dalam format yang sesuai untuk proses pemodelan K-Means.

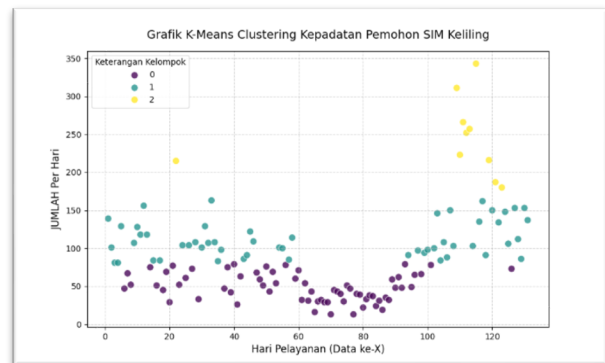
Proses persiapan tersebut menghasilkan dataset yang lebih terstruktur sehingga dapat digunakan pada tahap modeling. Pemilihan data dilakukan dengan mempertahankan atribut yang berhubungan langsung dengan tujuan analisis tingkat kepadatan pelayanan.

Hasil Modeling K-Means

K-Means diterapkan dengan tiga kluster, yaitu Sepi, Sedang, dan Ramai. Pada perhitungan manual, centroid awal ditetapkan sebesar 40, 100, dan 200. Proses pengelompokan dilakukan berdasarkan jarak terdekat antara data jumlah pemohon dan centroid.

Pada hasil akhir perhitungan manual, terdapat 63 data pada kluster Sepi, 58 data pada kluster Sedang, dan 10 data pada kluster Ramai dari total 131 hari operasional. Dengan demikian, kelompok Sepi merupakan kelompok dengan jumlah hari terbanyak, sedangkan kelompok Ramai memiliki jumlah hari paling sedikit.

Distribusi tersebut menunjukkan bahwa sebagian besar hari pelayanan berada pada tingkat kepadatan Sepi sampai Sedang, sedangkan kondisi Ramai muncul pada jumlah hari yang lebih terbatas. Informasi tersebut menjadi dasar untuk menentukan prioritas dalam pengelolaan layanan.



Gambar 3. Visualisasi Scatter Plot Pembuktian Partisi Objek Berdasarkan Koordinat Sentroid Manual

Visualisasi scatter plot digunakan untuk memperlihatkan hasil partisi objek berdasarkan kedekatan data terhadap koordinat centroid pada proses K-Means. Setiap objek dikelompokkan ke dalam kluster berdasarkan jarak terdekat dengan centroid sehingga terbentuk tiga kelompok tingkat kepadatan, yaitu Sepi, Sedang, dan Ramai. Visualisasi ini memperlihatkan distribusi objek pada masing-masing kluster dan mendukung hasil perhitungan clustering secara manual.

Validasi Hasil Menggunakan RapidMiner dan Python

Untuk menguji konsistensi model, hasil perhitungan manual dibandingkan dengan hasil pemodelan menggunakan RapidMiner dan Python. RapidMiner menghasilkan 60 hari kategori Sepi, 54 hari Sedang, dan 17 hari Ramai. Sementara itu, Python menghasilkan 67 hari Sepi, 54 hari Sedang, dan 10 hari Ramai.

Tabel 2. Perbandingan Distribusi Hasil Clustering

Kategori Kepadatan	K-Means Manual	Komputasi RapidMiner	Komputasi Python
Sepi	63 Hari	60 Hari	67 Hari
Sedang	58 Hari	54 Hari	54 Hari
Ramai	10 Hari	17 Hari	10 Hari
Total Data	131 Hari	131 Hari	131 Hari

Perbedaan jumlah anggota pada masing-masing perangkat lunak menunjukkan adanya deviasi pada hasil pengelompokan. Meskipun demikian, ketiga pendekatan tetap menghasilkan tiga kelompok dengan karakteristik yang sama, yaitu Sepi, Sedang, dan Ramai. Hasil pengujian silang menunjukkan tingkat keselarasan sebesar 91,60%

untuk RapidMiner dan 95,42% untuk Python terhadap hasil perhitungan manual.

Evaluasi Model

Evaluasi menggunakan Davies-Bouldin Index menunjukkan nilai 0,583 pada RapidMiner dan 0,520 pada Python. Nilai Python yang lebih rendah menunjukkan hasil pengelompokan yang lebih optimal berdasarkan indikator DBI yang digunakan dalam penelitian.

Tabel 3. Perbandingan Hasil Nilai Indeks Evaluasi K-Means

Perangkat Lunak / Tools	Sebaran Anggota Klaster (Sepi, Sedang, Ramai)	Nilai Davies-Bouldin Index (DBI)
Rapidminer	60 Hari, 54 Hari, 17 Hari	0.583
Python	67 Hari, 54 Hari, 10 Hari	0.520

Hasil tersebut menunjukkan bahwa model dapat digunakan untuk menggambarkan tingkat kepadatan pelayanan berdasarkan data historis. Pengujian menggunakan dua perangkat juga memberikan informasi mengenai konsistensi hasil pemodelan sebelum hasil tersebut digunakan pada tahap deployment.

Deployment dan Rekomendasi Penjadwalan

Tahap deployment merupakan tahap penerapan hasil analisis menjadi rekomendasi operasional. Hasil clustering menunjukkan bahwa setiap kategori memiliki karakteristik kepadatan yang berbeda sehingga dapat digunakan sebagai dasar pengambilan keputusan.

Klaster Ramai menjadi prioritas karena menunjukkan kondisi dengan jumlah pemohon yang lebih tinggi. Berdasarkan skripsi, kondisi tersebut dapat menjadi dasar untuk meningkatkan kesiapan sumber daya pelayanan, termasuk pengaturan armada, petugas, dan ketersediaan blanko SIM. Sebaliknya, klaster Sepi dapat digunakan untuk melakukan efisiensi penempatan sumber daya serta menjadi pertimbangan dalam kegiatan pemeliharaan teknis.

Klaster Sedang dapat diperlakukan sebagai kondisi pelayanan normal. Dengan demikian, hasil clustering memberikan tiga kondisi operasional yang dapat dijadikan dasar penyesuaian layanan, yaitu:

Tabel 4. Interpretasi Hasil Clustering

Klaster	Kondisi	Rekomendasi Penjadwalan
Sepi	Beban pelayanan rendah	Efisiensi sumber daya dan dapat dimanfaatkan untuk pemeliharaan
Sedang	Beban pelayanan normal	Mempertahankan pelayanan sesuai kondisi operasional
Ramai	Beban pelayanan tinggi	Menjadi prioritas dalam kesiapan armada, petugas, dan blanko

Rekomendasi tersebut tidak dimaksudkan sebagai pengganti keputusan operasional, tetapi sebagai informasi pendukung agar penjadwalan dapat mempertimbangkan pola historis jumlah pemohon. Skripsi juga mengembangkan hasil analisis ke dalam dasbor berbasis web sebagai bentuk implementasi hasil penelitian.

Penerapan hasil clustering memungkinkan pihak Satlantas melihat pola kepadatan berdasarkan data historis sehingga

penjadwalan tidak hanya bergantung pada pola tetap. Dengan pembaruan data secara berkala, hasil clustering dapat digunakan kembali untuk menyesuaikan rekomendasi terhadap perubahan pola jumlah pemohon.

D. PENUTUP

Simpulan

Berdasarkan hasil penelitian, penerapan metode CRISP-DM dan algoritma K-Means berhasil digunakan untuk menganalisis tingkat kepadatan pelayanan SIM Keliling di wilayah Polres Metro Depok. Data sebanyak 11.756 pemohon pada 131 hari operasional periode Januari sampai Juni 2025 dapat dikelompokkan menjadi tiga kategori, yaitu Sepi, Sedang, dan Ramai.

Hasil K-Means secara manual menghasilkan 63 hari kategori Sepi, 58 hari Sedang, dan 10 hari Ramai. Pengujian menggunakan RapidMiner menghasilkan 60 hari Sepi, 54 hari Sedang, dan 17 hari Ramai, sedangkan Python menghasilkan 67 hari Sepi, 54 hari Sedang, dan 10 hari Ramai. Evaluasi menghasilkan nilai Davies-Bouldin Index sebesar 0,583 pada RapidMiner dan 0,520 pada Python. Tingkat keselarasan terhadap perhitungan manual mencapai 91,60% pada RapidMiner dan 95,42% pada Python.

Hasil clustering dapat digunakan sebagai dasar pendukung dalam penyusunan rekomendasi penjadwalan layanan SIM Keliling. Klaster Ramai menjadi prioritas dalam kesiapan sumber daya, klaster Sedang menjadi kondisi pelayanan normal, sedangkan klaster Sepi dapat digunakan untuk meningkatkan efisiensi operasional. Dengan demikian, CRISP-DM dapat menjadi kerangka sistematis untuk mengubah data historis pemohon menjadi informasi yang mendukung pengambilan keputusan operasional.

Saran

Satlantas Polres Metro Depok disarankan melakukan pembaruan data pemohon secara berkala agar hasil pengelompokan tetap sesuai dengan perkembangan pola pelayanan. Selain itu, hasil analisis dapat digunakan bersama dengan dasbor yang telah dikembangkan sebagai pendukung pengambilan keputusan dalam penjadwalan layanan SIM Keliling. Penelitian selanjutnya dapat mengembangkan analisis dengan menggunakan data yang lebih panjang dan mempertimbangkan faktor lain yang relevan terhadap perubahan jumlah pemohon.

E. DAFTAR PUSTAKA

- Bakari, A. F., Rachman, E., & Katili, A. Y. (2024). Efektivitas pelayanan Surat Izin Mengemudi (SIM) mobil keliling oleh Polda Gorontalo. *Innovative: Journal of Social Science Research*, 4(2), 9509–9515. <https://doi.org/10.31004/innovative.v4i2.18973>
- Butsianto, S., & Saepudin, N. (2020). Penerapan data mining terhadap minat siswa dalam mata pelajaran matematika dengan metode K-Means. *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, 3(1), 51–59. <https://doi.org/10.32672/jnkti.v3i1.2008>

- Dimas Sulistiyo, R., & Shihab, M. R. (2023). Transformasi digital dalam pelayanan Surat Izin Mengemudi (SIM): Studi kasus Korlantas Polri. *Technomedia Journal*, 8(2SP), 189–204. <https://doi.org/10.33050/tmj.v8i2sp.2064>
- Mochammad, A., Dinata, Z. M., & Wada, I. A. (2024). Relevansi integritas dalam penerbitan SIM: Administrasi prosedural sebagai pendorong popularitas calo. *Inicio Legis*, 5(2), 1–15. <https://doi.org/10.21107/il.v5i2.26606>
- Nursia, A. N., Ramdhan, W., & Kifti, W. M. (2022). Analisis kelayakan penerima bantuan Covid-19 menggunakan metode K-Means. *Building of Informatics, Technology and Science (BITS)*, 3(4), 574–583. <https://doi.org/10.47065/bits.v3i4.1399>
- Puspitasari, C. (2022). *CRISP-DM (Cross-Industry Standard Process for Data Mining)*. Bina Nusantara University. <https://binus.ac.id/malang/2022/05/crisp-dm-cross-industry-standard-process-for-data-mining/>
- Rahmayani, S., Sumarno, S., & Siregar, Z. A. (2022). Analysis of K-Means algorithm for clustering of Covid-19 social assistance recipients. *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, 1(1), 77–84. <https://doi.org/10.55123/jomlai.v1i1.166>
- Ruswanti, D., Susilo, D., & Riani, R. (2024). Implementasi CRISP-DM pada data mining untuk melakukan prediksi pendapatan dengan algoritma C4.5. *Go Infotech: Jurnal Ilmiah STMIK AUB*, 30(1), 111–121. <https://doi.org/10.36309/goi.v30i1.266144>
- Salsabila, F., Fitrianti, I., Umaidah, Y., & Heryana, N. (2023). Penerapan metode CRISP-DM untuk analisa pendapatan bersih bulanan pekerja informal di Provinsi Jawa Barat dengan algoritma K-Means. *Dinamik*, 28(2), 97–104. <https://doi.org/10.35315/dinamik.v28i2.9454>
- Santiastry, S., Asriyanik, A., & Apriandari, W. (2024). Penerapan algoritma Naive Bayes dan metode CRISP-DM dalam prediksi hasil tes kemampuan bahasa Inggris mahasiswa. *JATI*, 8(5), 10432–10439. <https://doi.org/10.36040/jati.v8i5.11069>