

FRAMEWORK DATA MINING : SEBUAH SURVEI

Arya Ardhi Baskara, Nurul Maharani Piranti, Muhammad Fahrury Romdendine

Manajemen Teknologi Keimigrasian, Politeknik Imigrasi

Jalan Raya Gandul No. 4, Kota Depok, Indonesia

aryaaardhibaskara@gmail.com

ABSTRAK

Perkembangan pesat dalam ranah teknologi informasi telah meningkatkan kebutuhan akan metode data mining untuk menganalisis dan mengolah data dalam jumlah besar. Berbagai metodologi telah dikembangkan untuk mendukung proses ini, di antaranya Knowledge Discovery in Databases (KDD), Cross-Industry Standard Process for Data Mining (CRISP-DM), dan Sample, Explore, Modify, Model, and Assess (SEMMA). Penelitian ini bertujuan untuk mengevaluasi popularitas dan efektivitas masing-masing metodologi melalui pendekatan Systematic Literature Review berbasis PRISMA. Sebanyak 52 artikel dari tahun 2021 hingga 2025 dianalisis guna mengidentifikasi tren penggunaan metodologi dalam berbagai bidang, termasuk kesehatan, bisnis, teknologi, dan pendidikan. Hasil studi menunjukkan bahwa CRISP-DM adalah metodologi yang paling sering diterapkan karena fleksibilitasnya dalam berbagai sektor. Sementara itu, KDD dan SEMMA lebih banyak digunakan dalam konteks yang lebih spesifik. Studi ini menyoroti pentingnya pemilihan metodologi yang sesuai untuk memastikan efektivitas ekstraksi informasi dari data. Temuan penelitian ini diharapkan dapat menjadi referensi bagi akademisi, praktisi, dan peneliti dalam menentukan metodologi yang paling relevan berdasarkan karakteristik data dan tujuan analisis.

Kata kunci : Data Mining, KDD, CRISP-DM, SEMMA

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah membawa perubahan signifikan dalam berbagai aspek kehidupan, termasuk dalam dunia bisnis, pendidikan, kesehatan, serta pemerintahan. Teknologi informasi memungkinkan organisasi untuk mengelola data dalam jumlah besar secara lebih efisien, cepat, dan akurat, sehingga dapat mendukung proses pengambilan keputusan yang lebih tepat sasaran dan berbasis data. Digitalisasi di berbagai sektor menyebabkan volume data yang dihasilkan meningkat secara eksponensial setiap harinya, baik dari transaksi digital, interaksi pengguna, maupun dari sensor dan perangkat IoT (Internet of Things) [1].

Fenomena ini mendorong munculnya tantangan baru terkait bagaimana organisasi mampu mengelola data yang kompleks, menganalisisnya dengan pendekatan yang tepat, serta mengekstrak informasi yang memiliki nilai strategis. Data yang tidak dikelola dengan baik dapat menjadi beban, namun jika diproses secara efektif, dapat menjadi aset berharga dalam menciptakan keunggulan kompetitif. Oleh karena itu, kebutuhan akan metode seperti data mining menjadi semakin relevan dan krusial dalam era modern saat ini.

Perkembangan teknologi informasi telah membawa dunia memasuki era Big Data, di mana volume, variasi, dan kecepatan data yang dihasilkan terus meningkat secara eksponensial. Era Big Data ditandai oleh fenomena "5V" yaitu volume, variety, velocity, veracity, dan value yang menuntut organisasi untuk mampu mengolah data dalam skala besar dan beragam secara cepat dan akurat. Tantangan dalam era ini tidak hanya terletak pada penyimpanan data yang masif, tetapi juga bagaimana mengekstraksi informasi yang bernilai guna mendukung pengambilan

keputusan yang berbasis data (data-driven decision making) [2].

Oleh karena itu, kebutuhan akan metode analisis data yang sistematis dan efisien seperti data mining menjadi sangat mendesak di berbagai bidang.

Salah satu pendekatan terkait dengan tantangan data adalah data mining. Data mining atau penambangan data merupakan proses yang dilakukan untuk menemukan hubungan, pola, dan tren yang bermanfaat dengan mengeksplorasi kumpulan data dalam skala besar [3].

Berbagai teknik dan algoritma dalam data mining telah dikembangkan dan masing-masing memiliki keunggulan dan keterbatasan. Clustering digunakan untuk mengelompokkan data berdasarkan tingkat kesamaan, sedangkan Klasifikasi bertujuan untuk mengkategorikan data ke dalam kelas yang telah ditentukan sebelumnya. Regresi berfungsi memodelkan hubungan antar variabel guna memprediksi nilai numerik, sementara Asosiasi berfokus pada identifikasi hubungan antara item dalam suatu dataset. Pemilihan metode yang sesuai menjadi faktor krusial, karena efektivitasnya bergantung pada karakteristik data serta tujuan analisis yang ingin dicapai. Swamynathan, dalam penelitiannya, mengidentifikasi tiga metodologi utama dalam data mining, yaitu Knowledge Discovery in Databases (KDD), Cross-Industry Standard Process for Data Mining (CRISP-DM), serta Sample, Explore, Modify, Model, and Assess (SEMMA) [4].

Beberapa penelitian terdahulu menggunakan beragam framework data mining. Penelitian yang dilakukan oleh Mai Monir Ghazal Ahmed Hammad pada tahun 2022 berfokus pada penerapan proses KDD untuk memprediksi cost overrun dalam proyek

konstruksi. Penelitian ini memanfaatkan data mining sebagai bagian dari KDD untuk menggali pola dari data historis proyek, mengidentifikasi 12 faktor utama penyebab cost overrun [5].

Penelitian berikutnya oleh Puspitasari dkk pada tahun 2024 yang mengimplementasikan tahapan KDD (Knowledge Discovery in Databases) untuk menganalisis dan mengklasifikasikan risiko penyakit jantung menggunakan algoritma K-Nearest Neighbor (KNN) [6].

Hasanah dkk dalam penelitiannya pada tahun 2021 memprediksi curah hujan yang berpotensi menyebabkan banjir di Indonesia dengan memanfaatkan metodologi data mining CRISP-DM dan algoritma CART (Classification and Regression Tree) sebagai metode klasifikasinya [7].

Kemudian penelitian oleh Murti dkk pada tahun 2024 memprediksi jumlah kasus Covid-19 di Kabupaten Boyolali dengan membandingkan kinerja beberapa algoritma prediksi berbasis data mining [8].

Suwitono pada tahun 2022 mengimplementasikan algoritma Convolutional Neural Network (CNN) untuk klasifikasi citra daun menggunakan pendekatan data mining berbasis metode SEMMA [9]. Di tahun yang sama, Hasanati membahas analisis sentimen topik vaksin Covid-19 di Twitter dengan menggunakan algoritma Support Vector Machine (SVM) yang dipadukan dengan pendekatan lexicon-based [10].

Berangkat dari penelitian yang sudah dilakukan sebelumnya, penelitian ini bertujuan untuk mengeksplorasi dan menganalisis tren penggunaan berbagai framework yang diterapkan dalam penelitian data mining, dengan fokus utama pada perbandingan antara tiga framework populer, yaitu Knowledge Discovery in Databases (KDD), Cross Industry Standard Process for Data Mining (CRISP-DM), dan Sample, Explore, Modify, Model, and Assess (SEMMA). Studi ini dirancang untuk mengidentifikasi kelebihan, kekurangan, serta kecenderungan para peneliti dalam memilih dan mengimplementasikan masing-masing framework dalam proses eksekusi dataset pada berbagai bidang studi. Selain itu, penelitian ini juga akan mengevaluasi bagaimana framework tersebut diadopsi dalam konteks yang berbeda, seperti industri, akademik, maupun sektor pemerintahan, serta dalam berbagai jenis permasalahan data mining seperti klasifikasi, klustering, prediksi, dan asosiasi. Dengan menganalisis pola dan preferensi penggunaan framework tersebut, diharapkan hasil penelitian ini dapat memberikan kontribusi yang bermakna dalam memperkaya wawasan dan membantu peneliti maupun praktisi dalam menentukan framework yang paling sesuai dengan kebutuhan dan karakteristik penelitian data mining yang mereka lakukan.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah proses yang memanfaatkan teknik statistik, kecerdasan buatan, dan machine learning untuk mengekstraksi pengetahuan atau informasi tersembunyi dari kumpulan data dalam jumlah besar. Tujuan utama dari data mining adalah untuk menemukan pola-pola penting yang dapat mendukung proses pengambilan keputusan dalam organisasi maupun bisnis. Data mining hadir sebagai solusi untuk mengolah tumpukan data yang seringkali hanya menjadi arsip tanpa makna menjadi sumber informasi yang bernilai guna. Kondisi ini dikenal sebagai fenomena "rich of data but poor of information", di mana banyak organisasi yang memiliki data dalam jumlah besar, namun belum mampu mengubahnya menjadi insight yang bermanfaat [11].

Data mining merupakan proses iteratif yang secara otomatis dapat menemukan pola dan hubungan tersembunyi dalam data dengan tujuan memberikan indikasi atau prediksi yang berguna untuk pengambilan keputusan. Data mining tidak hanya memproses data yang sangat besar (big data) namun juga harus mampu memberikan hasil berupa pengetahuan atau insight yang mudah dipahami oleh pengguna. Hal ini menjadikan data mining sebagai alat penting dalam mendukung efisiensi operasional dan meningkatkan keunggulan kompetitif organisasi [12].

Manfaat data mining tidak hanya dirasakan di dunia bisnis tetapi juga di dunia akademis dan penelitian. Dalam dunia bisnis, data mining sering digunakan untuk analisis perilaku konsumen, deteksi fraud, hingga perencanaan kampanye pemasaran yang lebih terarah. Sedangkan di bidang ilmiah, data mining membantu dalam menganalisis data yang bersumber dari eksperimen, observasi, atau sensor yang menghasilkan data dalam jumlah besar. Dengan demikian, data mining menjadi salah satu teknologi penting yang digunakan untuk mengoptimalkan informasi yang ada agar dapat mendukung proses pengambilan keputusan yang berbasis data [13].

2.2. Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) merupakan proses penting dalam data mining yang bertujuan untuk mengekstraksi pola dan pengetahuan yang bermakna dari sejumlah besar data yang kompleks. KDD sangat relevan dalam era digital saat ini. Aktivitas manusia umumnya meninggalkan jejak data dalam jumlah besar, seperti check-in media sosial, ulasan daring, dan data perjalanan. KDD berperan untuk memproses dan menganalisis data ini guna membangun pemahaman yang lebih dalam mengenai aktivitas spasio-temporal manusia, yang kemudian digunakan dalam berbagai aplikasi seperti perencanaan kota, pengelolaan lalu lintas, hingga mitigasi bencana. Dalam konteks ini, KDD tidak hanya mencakup teknik pattern discovery, tetapi juga model prediktif dan embedding multimodal yang

mampu memetakan korelasi lintas-modality secara langsung untuk menghasilkan pengetahuan yang actionable [14].

KDD berkembang dari kebutuhan untuk menganalisis big data yang semakin kompleks dan tidak dapat lagi diselesaikan dengan metode statistik konvensional. KDD menggabungkan prinsip-prinsip dari statistik, kecerdasan buatan, dan sistem basis data untuk menemukan pola yang valid, bermakna, dan dapat diinterpretasikan dalam data. Lebih jauh lagi, KDD tidak hanya memperkuat proses prediktif dalam data mining, tetapi juga mendukung penemuan kausalitas yang kompleks dengan memanfaatkan algoritma machine learning seperti decision tree dan random forest yang mampu mengeksplorasi heterogenitas kausal dalam data. Dalam kerangka computational social science, KDD menjadi pendekatan yang integral dalam memahami perilaku individu dan kolektif serta dinamika sosial yang kompleks di era big data [15].

KDD berfungsi sebagai kerangka kerja sistematis yang menjembatani proses eksploratif dan konfirmatori dalam penemuan pengetahuan berbasis data. Dari sisi teknis, KDD melalui tahap-tahap seperti pemahaman domain, pemilihan dataset, pra-pemrosesan data, reduksi data, pemilihan metode dan algoritma data mining, hingga interpretasi hasil. Proses ini selaras dengan pendekatan dialektis yang mengintegrasikan metode induktif dan deduktif, memungkinkan peneliti untuk tidak hanya menemukan pola yang tersembunyi tetapi juga memvalidasi temuan secara teoritis dan praktis.

2.3. Cross-Industry Standard Process for Data Mining (CRISP-DM)

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah metodologi yang telah menjadi standar industri dalam pelaksanaan proyek data mining lintas berbagai sektor. CRISP-DM membagi proses data mining menjadi enam tahap utama yang iteratif, yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. CRISP-DM sangat berguna dalam proyek data mining berbasis kecerdasan buatan karena pendekatan siklus hidupnya yang terstruktur dan fleksibel, memfasilitasi pemahaman bisnis serta eksplorasi data yang lebih terarah. Dalam penelitian mereka mengenai prediksi kinerja siswa di Kolombia menggunakan CRISP-DM, proses ini terbukti mampu memberikan landasan sistematis untuk memilih atribut yang relevan dan mengevaluasi berbagai model pembelajaran mesin secara efektif dalam konteks pendidikan [16].

Di sisi lain, studi oleh Brzozowska dkk [ada tahun 2023 menyoroti pentingnya penerapan CRISP-DM dalam lingkungan manufaktur, khususnya untuk mendukung pengambilan keputusan di bidang perakitan produksi. Penelitian ini menunjukkan bagaimana tiga fase awal CRISP-DM (business understanding, data understanding, dan data preparation) digunakan untuk mempersiapkan data

yang berkualitas sebelum proses modeling dilakukan. Selain itu, penelitian ini juga mempertegas bahwa fase data preparation sering kali menjadi tahap paling memakan waktu dan sumber daya, terutama saat menangani data industri yang kompleks dan tidak terstruktur. CRISP-DM membantu perusahaan dalam membangun model prediksi yang dapat digunakan untuk memperkirakan waktu siklus perakitan produk, sehingga mampu meningkatkan efisiensi dalam proses bisnis dan perencanaan produksi [17].

Secara konseptual, CRISP-DM juga dipandang sebagai kerangka desain model dalam konteks teori desain model. CRISP-DM tidak hanya berfungsi sebagai proses teknis dalam data mining, tetapi juga dapat dianalisis sebagai pendekatan desain model yang sistematis. Dalam perspektif ini, CRISP-DM menampilkan karakteristik yang serupa dengan proses desain di berbagai domain lain seperti sistem rekayasa dan perangkat lunak. Penelitian mereka menunjukkan bahwa CRISP-DM sangat berfokus pada solusi (solution-oriented) dibandingkan dengan pendekatan desain lain yang lebih problem-oriented. Selain itu, CRISP-DM memiliki pola iterasi yang unik dalam mendesain model data mining, dengan penekanan pada evaluasi dan penyempurnaan model berdasarkan umpan balik dari proses sebelumnya [18].

2.4. Sample, Explore, Modify, Model, and Assess (SEMMA)

SEMMA (Sample, Explore, Modify, Model, and Assess) adalah metodologi data mining yang dikembangkan oleh SAS Institute untuk memberikan pendekatan sistematis dalam eksplorasi dan analisis data. SEMMA berfokus pada eksplorasi data yang ekstensif sebelum melakukan pemodelan, sehingga memungkinkan pemahaman yang lebih baik terhadap pola dan tren dalam dataset. Tahapan dalam SEMMA diawali dengan pengambilan sampel data, eksplorasi untuk menemukan pola awal, modifikasi data agar sesuai untuk analisis lebih lanjut, pemodelan menggunakan teknik statistik atau machine learning, dan terakhir penilaian hasil untuk memastikan model yang dibangun sesuai dengan tujuan analisis [19].

Dalam konteks data mining, SEMMA sering dibandingkan dengan metodologi lain seperti CRISP-DM. Perbedaan utama antara keduanya terletak pada struktur pendekatannya: CRISP-DM menekankan pemahaman bisnis sebelum masuk ke eksplorasi data, sedangkan SEMMA lebih langsung ke dalam proses eksplorasi dan pemodelan data. Pendekatan SEMMA ini memungkinkan fleksibilitas dalam eksplorasi data, tetapi bisa kurang terstruktur dalam menentukan tujuan bisnis dan strategi pengolahan data yang lebih luas. Oleh karena itu, dalam beberapa kasus, kombinasi SEMMA dan CRISP-DM digunakan untuk mengatasi keterbatasan masing-masing model dan meningkatkan efektivitas proyek data mining [20].

Penerapan SEMMA dalam data mining telah banyak digunakan di berbagai sektor, termasuk keuangan, pemasaran, dan ilmu kesehatan. Dengan

pendekatan berbasis eksplorasi data, SEMMA dapat membantu dalam mendeteksi anomali dalam data keuangan, mengidentifikasi segmentasi pelanggan yang lebih tepat dalam pemasaran, serta menganalisis pola dalam data medis untuk mendukung diagnosis dan perawatan pasien. Namun, meskipun SEMMA menawarkan struktur yang sistematis, efektivitasnya sangat tergantung pada kualitas data awal serta pemahaman yang baik terhadap domain bisnis yang dianalisis [21].

2.5. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) merupakan panduan pelaporan berbasis bukti yang dikembangkan untuk meningkatkan transparansi dan akurasi dalam menyusun laporan kajian sistematis dan meta-analisis. Awalnya dirancang untuk kebutuhan riset di bidang kesehatan, PRISMA kini telah berkembang dan digunakan secara luas di berbagai disiplin ilmu, termasuk manajemen, keuangan, dan teknik lingkungan. PRISMA menekankan pentingnya pelaporan yang lengkap dan sistematis melalui tahapan yang mencakup identifikasi, penyaringan, penilaian kelayakan, dan inklusi literatur yang relevan. Selain menyediakan checklist, PRISMA juga dilengkapi dengan diagram alur yang mempermudah peneliti dalam menelusuri dan melaporkan proses seleksi artikel secara transparan [22].

3. METODE PENELITIAN

Pedoman yang paling sering digunakan dalam pelaporan tinjauan sistematis, khususnya yang mencakup aspek pencarian literatur adalah Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Pedoman ini dirancang untuk memastikan bahwa proses pelaporan tinjauan sistematis dan meta-analisis dilakukan secara sistematis transparan, dan sesuai dengan standard yang telah ditetapkan, sehingga meningkatkan kualitas dan reproduksibilitas penelitian. Proses ini terdiri dari 5 tahap diantaranya menentukan persyaratan kelayakan, menentukan referensi informasi, menentukan literatur, mengumpulkan data, dan menentukan jenis data [23]

Tahapan penelitian menggunakan metode PRISMA diantaranya sebagai berikut

- a. Tahap 1: Menentukan Kriteria Kelayakan Artikel
Kriteria kelayakan ditentukan berdasarkan kriteria inklusi (IC), yaitu:
 - IC1: Artikel harus didasarkan pada penelitian asli yang telah dilakukan.
 - IC2: Artikel yang dipublikasikan antara tahun 2021 hingga 2025.
 - IC3: Artikel ini bertujuan untuk menganalisis metode yang digunakan oleh peneliti lain dalam melakukan proses data mining.
- b. Tahap 2: Menentukan Sumber Informasi
Literatur dapat ditemukan melalui basis data daring dalam berbagai koleksi akademik diantaranya

adalah Google Scholar, ScienceDirect, SpringerOpen, dan ProQuest.

c. Tahap 3: Seleksi Literatur

- Pemilihan kata kunci yang digunakan adalah “KDD”, “CRISP-DM”, dan “SEMMA”.
- Untuk meninjau dan memilih judul, ringkasan, serta artikel, kata kunci yang diperoleh dari hasil pencarian akan dievaluasi ulang dengan menerapkan tahapan 3 hingga 4 guna memastikan kesesuaiannya dengan penelitian ini.

d. Tahap 4: Pengumpulan Data

Tabel 1. Pengumpulan Data

Sumber	Studi Literatur yang ditemukan			Kandidat	Terpilih
	KDD	CRISP-DM	SEMMA		
Google Scholar	35.200	13.200	4.150	236	30
Science Direct	2.676	1.784	114	63	10
Springer Open	96	22	4	24	7
Pro Quest	16.954	1.586	3.769	58	5
Total	54.926	16.592	8.037	381	52

Data yang dikumpulkan dapat dilihat dalam Tabel 1. Data dikumpulkan melalui pembuatan formulir ekstraksi data manual. Sebanyak 1346 artikel dievaluasi menggunakan kata kunci “KDD”, “CRISP-DM”, dan “SEMMA” dari semua sumber dan kriteria yang tersedia. Dari jumlah tersebut, 381 artikel memenuhi persyaratan awal. Setelah proses penyaringan tambahan, terdapat 52 artikel dipilih untuk penelitian ini.

e. Tahap 5: Pemilihan Item Data

Data didapatkan dari artikel terpilih yang berisi satu dari framework data mining antara KDD, CRISP-DM, atau SEMMA.

4. HASIL DAN PEMBAHASAN

Penelitian ini mengkaji 3 metode dalam data mining yang digunakan oleh para peneliti di berbagai aspek. Metode tersebut diantaranya adalah KDD (Knowledge Discovery in Databases), CRISP-DM (Cross-Industry Standard Process for Data Mining), dan SEMMA (Sample, Explore, Modify, Model, dan Assess). Penelitian ini bertujuan untuk menyelidiki pendekatan yang paling banyak diimplementasikan oleh peneliti lain dalam menggunakan metodologi data mining. Selain itu, penelitian ini juga menganalisis tren literatur "Selected Studies" berdasarkan nama penulis, sumber publikasi, tahun publikasi, dan jenis literatur. Tabel 2 menunjukkan sumber publikasi yang digunakan.

Tabel 2. Sumber Publikasi

No	Judul	Tahun	Jenis
1	An optimized ... [24]	2021	Konferensi
2	Adapting ... [25]	2021	Jurnal
3	Komparasi ... [26]	2021	Jurnal
4	Adapting ...[27]	2021	Konferensi
5	Implementasi ...[7]	2021	Jurnal
6	Using data-mining ...[28]	2021	Konferensi
7	The use of ... [29]	2021	Jurnal
8	A CRISP-DM ... [30]	2021	Konferensi
9	A Methodology ... [31]	2021	Konferensi
10	Application ... [5]	2022	Jurnal
11	Optimization ... [32]	2022	Jurnal
12	Implementasi ... [33]	2022	Jurnal
13	Analisis ... [34]	2022	Jurnal
14	Penentuan ... [35]	2022	Konferensi
15	Penerapan ... [36]	2022	Jurnal
16	Cross-Industry ... [37]	2022	Konferensi
17	Penerapan ... [38]	2022	Jurnal
18	Applying ... [39]	2022	Jurnal
19	Predictive ... [16]	2022	Konferensi
20	Implementasi ... [9]	2022	Jurnal
21	Implementation ... [10]	2022	Konferensi
22	Human ... [40]	2022	Konferensi
23	Data mining ... [41]	2022	Konferensi
24	Improved ... [42]	2023	Jurnal
25	Knowledge ... [43]	2023	Jurnal
26	Penerapan ... [44]	2023	Jurnal
27	Penerapan ... [45]	2023	Jurnal
28	Analisis ... [46]	2023	Jurnal
29	Crop Prediction ... [47]	2023	Konferensi
30	Marketing ... [48]	2023	Jurnal
31	Using Supervised ...[49]	2023	Jurnal
32	Penerapan ... [50]	2023	Jurnal

No	Judul	Tahun	Jenis
33	Penerapan ... [51]	2023	Jurnal
34	Analysis ... [52]	2023	Jurnal
35	A clinical ... [53]	2023	Jurnal
36	Identification ... [54]	2023	Konferensi
37	Utilizing ... [55]	2024	Jurnal
38	Understanding ... [56]	2024	Jurnal
39	Application of ... [57]	2024	Jurnal
40	Heart Disease: ... [6]	2024	Jurnal
41	Realising ... [58]	2024	Jurnal
42	A CRISP-DM ... [59]	2024	Jurnal
43	Sentiment ... [60]	2024	Jurnal
44	Implementation ... [61]	2024	Jurnal
45	Analisis ... [8]	2024	Jurnal
46	Data mining ... [62]	2024	Jurnal
47	Harnessing ... [63]	2024	Konferensi
48	Textual ... [64]	2024	Jurnal
49	The RFMRv ... [65]	2024	Jurnal
50	Identifying ... [66]	2024	Jurnal
51	Hybrid ... [67]	2024	Jurnal
52	Optimized ... [68]	2024	Jurnal

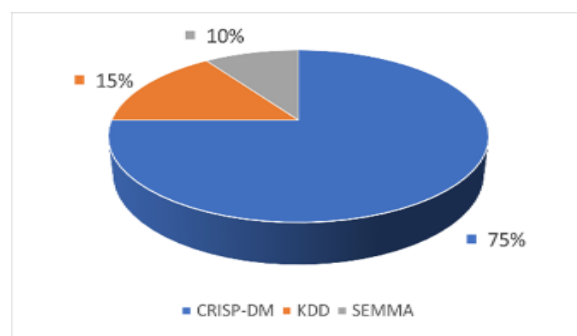
Tabel 2 menjelaskan tentang berbagai artikel dari rentang tahun 2021 hingga awal tahun 2025 yang menggunakan metode dalam data mining. Tabel meliputi nama penulis, tahun publikasi, metodologi, algoritma, teknik dan aspek yang dibahas dalam artikel. Analisis berdasarkan Tabel 3 yang ditampilkan dalam diagram di bawah ini menunjukkan bahwa metode CRISP-DM merupakan metode yang paling banyak digunakan dalam implementasi data mining. Metode ini memiliki banyak keunggulan karena dapat fleksibel dan dapat diterapkan di berbagai kebutuhan pengolahan data.

Tabel 3. Framework Data Mining

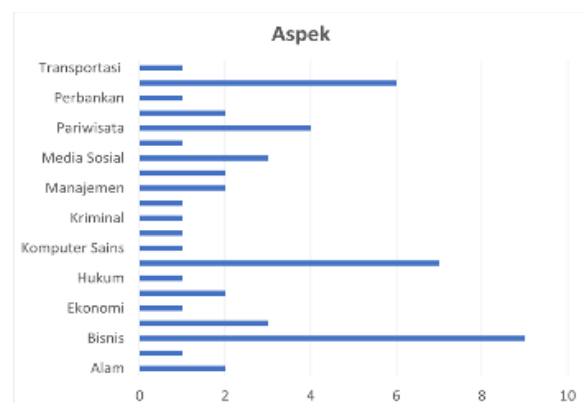
Penulis	Tahun	Frame work	Teknik	Aspek
Mavrogiorgou et al. [24]	2021	KDD	-	Kesehatan
Cazacu dan Titan [25]	2021	CRISP-DM	Regression	Teknologi
Navisa et al. [26]	2021	CRISP-DM	Classification	Musik
Plotnikova et al. [27]	2021	CRISP-DM	-	Finansial
Hasanah dan Soim [7]	2021	CRISP-DM	Classification	Alam
Gharbi et al. [28]	2021	CRISP-DM	Prediction	Teknologi
Arwa A. Jamjoom [29]	2021	CRISP-DM	Classification dan Regression	Bisnis
Cunha et al. [30]	2021	CRISP-DM	Prediction	Kesehatan
Costa dan Aparicio [31]	2021	SEMMA	-	Kesehatan
Ghazal dan Hammad [5]	2022	KDD	Clustering dan Classification	Data Sains
Kurniawan dan Yasir [32]	2022	CRISP-DM	Classification	Media Sosial
Dhewayani et al [33]	2022	CRISP-DM	Clustering	Alam
Singgale [34]	2022	CRISP-DM	Classification	Pariwisata
Shedriko dan Firdaus [35]	2022	CRISP-DM	Classification dan Prediction	Pendidikan
Wurijanto et al [36]	2022	CRISP-DM	Prediction	Perbankan
Rotty et al. [37]	2022	CRISP-DM	Regression	Bisnis
Muttaqin dan Hermanto [38]	2022	CRISP-DM	Clustering	Bisnis
Plotnikova et al [39]	2022	CRISP-DM	-	Finansial
Solano et al. [16]	2022	CRISP-DM	Prediction	Pendidikan
Suwitono dan Kaunang [9]	2022	SEMMA	Classification	Biologi
Hasanati et al. [10]	2022	SEMMA	Classification	Media Sosial
Rocha et al [40]	2022	SEMMA	Clustering dan Classification	Teknologi
Drowatzky et al [41]	2022	SEMMA	-	Teknologi
Alzyoud [42]	2023	KDD	-	Manajemen
Ortega-Guzmán [43]	2023	KDD	GraphMining	Data Sains

Penulis	Tahun	Frame work	Teknik	Aspek
Pambudi [44]	2023	CRISP-DM	Regression	Ekonomi
Singgalel [45]	2023	CRISP-DM	Classification	Pariwisata
Singgalel [46]	2023	CRISP-DM	SMOTE	Pariwisata
Rahmadi et al [47]	2023	CRISP-DM	-	Data Sains
Iswavigra et al [48]	2023	CRISP-DM	Clustering	Bisnis
Lee dan Baxter [49]	2023	CRISP-DM	Classification dan Prediction	Hukum
Sastya dan Nugraha [50]	2023	CRISP-DM	-	Bisnis
Singgalel [51]	2023	CRISP-DM	STP	Bisnis
Gupta dan Sharma [52]	2023	CRISP-DM	Regression	Kriminal
Bilal Mirza et al. [53]	2023	CRISP-DM	Prediction	Teknologi
Julian Weller [54]	2023	CRISP-DM	Regression	Manufaktur
Singgalel [55]	2024	KDD	Clustering	Bisnis
Singgalel [56]	2024	KDD	EDA dan NLP	Bisnis
Llatas Oliver et al. [57]	2024	KDD	Correlation Analysis	Konstruksi
Puspitasari et al. [6]	2024	KDD	Clustering	Kesehatan
Bokrantz dan Subramanian [58]	2024	CRISP-DM	Clustering, Classification dan Regression	Manufaktur
Bemthuis et al. [59]	2024	CRISP-DM	Process Mining	Komputer Sains
Singgalel [60]	2024	CRISP-DM	Classification	Bisnis
Singgalel [61]	2024	CRISP-DM	Classification	Pariwisata
Murti [8]	2024	CRISP-DM	Prediction	Kesehatan
Fernandes dan Celho [62]	2024	CRISP-DM	Prediction	Transportasi
Regina Sousa et al. [63]	2024	CRISP-DM	Prediction	Kesehatan
Latifa Sudar [64]	2024	CRISP-DM	Classification	Media Sosial
Mirfakhraei [65]	2024	CRISP-DM	Clustering	Manajemen
Mulyani Karmagatri [66]	2024	CRISP-DM	Classification	Lingkungan
Wu [67]	2024	CRISP-DM	Regression	Teknologi
Muningsih [68]	2024	CRISP-DM	Classification	Kesehatan

Pada tabel 3 menampilkan framework dan teknik data mining di berbagai aspek yang digunakan peneliti. Berdasarkan analisis yang telah dilakukan, aspek Bisnis jumlah terbanyak diantara aspek penelitian lainnya yaitu 9 artikel. Posisi berikutnya ditempati aspek Kesehatan sebanyak 7 artikel dan Teknologi sebanyak 6 artikel. Berikutnya ditempati aspek pariwisata sebanyak 4 artikel. Aspek Data Sains dan Media Sosial masing-masing terdapat 3 artikel. Sedangkan, aspek Alam, Finansial, Manajemen, Manufaktur, dan Pendidikan masing-masing sebanyak 2 artikel. Posisi terakhir ditempati berbagai aspek seperti Biologi, Ekonomi, Hukum, Komputer Sains, Konstruksi, Kriminal, Lingkungan, Musik, Perbankan, dan Transportasi masing-masing terdapat 1 peneliti. Berdasarkan data tersebut dapat disimpulkan bahwa penggunaan metedologi data mining dapat diterapkan di berbagai aspek permasalahan.



Gambar 1. Framework yang digunakan dalam penelitian



Gambar 2. Aspek penelitian yang dapat diimplementasikan dalam framework data mining

Gambar 1 menunjukkan diagram pie yang menggambarkan persentase penggunaan framework data mining. Framework CRISP-DM mendominasi dengan proporsi sebesar 75%, yang menunjukkan bahwa framework ini adalah yang paling umum digunakan dalam proses data mining. Disusul oleh KDD dengan 15% dan SEMMA dengan 10%. Hal ini menandakan bahwa CRISP-DM masih menjadi standar dalam banyak proyek data mining karena strukturnya yang fleksibel dan luas digunakan di industri maupun penelitian.

Gambar 2 menampilkan diagram batang horizontal yang menunjukkan berbagai aspek atau bidang di mana data mining dapat diimplementasikan. Dari gambar tersebut, terlihat bahwa aspek Bisnis menjadi bidang yang paling banyak

diimplementasikan dengan skor tertinggi, diikuti oleh Hukum dan Perbankan. Selain itu, bidang seperti Pariwisata, Media Sosial, dan Manajemen juga cukup sering menjadi area penerapan data mining. Sementara aspek seperti Kriminal, Transportasi, dan Ekonomi memiliki frekuensi implementasi yang lebih rendah.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil kajian literatur sistematis terhadap 52 artikel yang dipublikasikan antara tahun 2021 hingga 2025, dapat disimpulkan bahwa metode CRISP-DM merupakan framework data mining yang paling banyak digunakan oleh para peneliti dibandingkan dengan KDD dan SEMMA. CRISP-DM dinilai lebih fleksibel dan mampu diterapkan di berbagai bidang, mulai dari kesehatan, bisnis, teknologi, pendidikan, hingga pariwisata. Meskipun KDD dan SEMMA juga masih banyak diimplementasikan, penggunaannya lebih banyak terfokus pada bidang atau konteks tertentu. Temuan ini menunjukkan bahwa pemilihan framework yang tepat sangat bergantung pada tujuan penelitian dan karakteristik data yang digunakan.

Sebagai saran, peneliti dan praktisi diharapkan dapat mempertimbangkan CRISP-DM sebagai pilihan utama ketika menghadapi proyek data mining yang memerlukan fleksibilitas tinggi dan proses yang terstruktur. Namun demikian, jika penelitian atau proyek memiliki kebutuhan spesifik yang lebih cocok dengan karakteristik KDD atau SEMMA, maka kedua framework tersebut tetap relevan untuk diadopsi. Selain itu, diperlukan pemahaman yang lebih dalam terhadap masing-masing tahapan dalam framework agar proses ekstraksi pengetahuan berjalan lebih optimal dan memberikan hasil yang akurat serta bermanfaat bagi pengambilan keputusan.

DAFTAR PUSTAKA

- [1] J. P. Laudon, K. C., & Laudon, *Management Information Systems: Managing the Digital Firm*. Pearson, 2020.
- [2] M. Chen, S. Mao, and Y. Liu, "Big Data: A Survey," *Mob. Networks Appl.*, vol. 19, no. 2, pp. 171–209, Apr. 2014, doi: 10.1007/s11036-013-0489-0.
- [3] D. A. Manalu and G. Gunadi, "IMPLEMENTASI METODE DATA MINING K-MEANS CLUSTERING TERHADAP DATA PEMBAYARAN TRANSAKSI MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON PADA CV DIGITAL DIMENSI," *Infotech J. Technol. Inf.*, vol. 8, no. 1, pp. 43–54, Jun. 2022, doi: 10.37365/jti.v8i1.131.
- [4] M. Swamynathan, *Mastering Machine Learning with Python in Six Steps*. Berkeley, CA: Apress, 2017. doi: 10.1007/978-1-4842-2866-1.
- [5] M. M. Ghazal and A. Hammad, "Application of knowledge discovery in database (KDD) techniques in cost overrun of construction projects," *Int. J. Constr. Manag.*, vol. 22, no. 9, pp. 1632–1646, Jul. 2022, doi: 10.1080/15623599.2020.1738205.
- [6] D. Puspitasari, A. J. Aprian, E. D. Sikumbang, K. Ramanda, S. H. Sukmana, and Q. N. Azizah, "Heart Disease: Application of the K-Nearest Neighbor (KNN) Method," *Ingénierie des systèmes d'Inf.*, vol. 29, no. 4, pp. 1275–1281, Aug. 2024, doi: 10.18280/isi.290403.
- [7] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [8] B. Juli and E. Wisnu Murti, "JIKES : JURNAL ILMU KESEHATAN Analisis Dan Perbandingan Algoritma Prediksi Dalam Mengetahui Perkiraan Peningkatan Jumlah Kasus COVID-19 Di Kabupaten Boyolali Dengan Metodologi CRISP-DM," vol. 3, pp. 24–34, 2024.
- [9] Y. A. Suwitono and F. J. Kaunang, "Implementasi Algoritma Convolutional Neural Network (CNN) Untuk Klasifikasi Daun Dengan Metode Data Mining SEMMA Menggunakan Keras," *J. Komtika (Komputasi dan Inform.)*, vol. 6, no. 2, pp. 109–121, 2022, doi: 10.31603/komtika.v6i2.8054.
- [10] N. Hasanati, Q. Aini, and A. Nuri, "Implementation of Support Vector Machine with Lexicon Based for Sentiment Analysis on Twitter," *2022 10th Int. Conf. Cyber IT Serv. Manag. CITSM 2022*, 2022, doi: 10.1109/CITSM56380.2022.9935887.
- [11] J. Rady, A. Jr, and E. Polit, *Data Warehouse – Data Mining*, no. March 2021. 2001.
- [12] D. Listriani and A. H. Setyaningrum, "Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Pola Belanja Konsumen (Studi Kasus Toko Buku Gramedia Bintaro)," *Int. J. Sci. Eng. Res. (IJOSER)*, vol. 3, no. 4, p. 2, 2015, [Online]. Available: <http://journal.uinjkt.ac.id/index.php/ti/article/view/5602/3619>
- [13] A. B. Berek, S. M. Isa, and U. B. Nusantara, "Data Mining untuk Memprediksi Animo Masyarakat terhadap Proses Penerimaan Peserta Didik Baru," vol. 6, no. 1, pp. 219–226, 2025.
- [14] C. Zhang and J. Han, *Data Mining and Knowledge Discovery*. Springer Singapore, 2021. doi: 10.1007/978-981-15-8983-6_42.
- [15] X. Shu and Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning," *Soc. Sci. Res.*, vol. 110, no. April 2022, p. 102817, 2023, doi: 10.1016/j.ssresearch.2022.102817.
- [16] J. A. Solano, D. J. Lancheros Cuesta, S. F. Umaña Ibáñez, and J. R. Coronado-Hernández, "Predictive models assessment based on CRISP-

- DM methodology for students performance in Colombia - Saber 11 Test,” *Procedia Comput. Sci.*, vol. 198, pp. 512–517, 2022, doi: 10.1016/j.procs.2021.12.278.
- [17] J. Brzozowska, J. Pizoń, G. Baytikenova, A. Gola, A. Zakimova, and K. Piotrowska, “Data Engineering in Crisp-Dm Process Production Data – Case Study,” *Appl. Comput. Sci.*, vol. 19, no. 3, pp. 83–95, 2023, doi: 10.35784/acs-2023-26.
- [18] U. Kannengiesser and J. S. Gero, “Modelling the Design of Models: an Example Using Crisp-Dm,” *Proc. Des. Soc.*, vol. 3, no. July, pp. 2705–2714, 2023, doi: 10.1017/pds.2023.271.
- [19] H. J. G. Palacios, R. A. J. Toledo, G. A. H. Pantoja, and Á. A. M. Navarro, “A comparative between CRISP-DM and SEMMA through the construction of a MODIS repository for studies of land use and cover change,” *Adv. Sci. Technol. Eng. Syst.*, vol. 2, no. 3, pp. 598–604, 2017, doi: 10.25046/aj020376.
- [20] U. Shafique and H. Qaiser, “A Comparative Study of Data Mining Process Models (KDD , CRISP-DM and SEMMA),” *Int. J. Innov. Sci. Res.*, vol. 12, no. 1, pp. 217–222, 2014, [Online]. Available: <http://www.ijisr.issr-journals.org/>
- [21] Omari Firas, “A combination of SEMMA & CRISP-DM models for effectively handling big data using formal concept analysis based knowledge discovery: A data mining approach,” *World J. Adv. Eng. Technol. Sci.*, vol. 8, no. 1, pp. 009–014, 2023, doi: 10.30574/wjaets.2023.8.1.0147.
- [22] R. Sarkis-Onofre, F. Catalá-López, E. Aromataris, and C. Lockwood, “How to properly use the PRISMA Statement,” *Syst. Rev.*, vol. 10, no. 1, pp. 13–15, 2021, doi: 10.1186/s13643-021-01671-z.
- [23] A. Liberati *et al.*, “The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration,” *J. Clin. Epidemiol.*, vol. 62, no. 10, pp. e1–e34, Oct. 2009, doi: 10.1016/j.jclinepi.2009.06.006.
- [24] A. Mavrogiorgou, A. Kiourtis, G. Manias, and D. Kyriazis, “An Optimized KDD Process for Collecting and Processing Ingested and Streaming Healthcare Data,” in *2021 12th International Conference on Information and Communication Systems (ICICS)*, IEEE, May 2021, pp. 49–56. doi: 10.1109/ICICS52457.2021.9464551.
- [25] M. Cazacu and E. Titan, “Adapting CRISP-DM for Social Sciences,” *BRAIN. Broad Res. Artif. Intell. Neurosci.*, vol. 11, no. 2sup1, pp. 99–106, 2020, doi: 10.18662/brain/11.2sup1/97.
- [26] S. Navisa, Luqman Hakim, and Aulia Nabilah, “Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM,” *J. Sist. Cerdas*, vol. 4, no. 2, pp. 114–125, 2021, doi: 10.37396/jsc.v4i2.162.
- [27] V. Plotnikova, M. Dumas, and F. Milani, “Adapting the CRISP-DM Data Mining Process: A Case Study in the Financial Services Domain,” *Lect. Notes Bus. Inf. Process.*, vol. 415 LNBP, pp. 55–71, 2021, doi: 10.1007/978-3-030-75018-3_4.
- [28] S. Al Gharbi, A. Al-Majed, A. Abdulraheem, S. Patil, and S. Elkhatatny, “Using data-mining CRISP-DM methodology to predict drilling troubles in real-time,” *Proc. IADC/SPE Asia Pacific Drill. Technol. Conf. APDT*, 2021, doi: 10.2118/201026-MS.
- [29] A. A. Jamjoom, “The use of knowledge extraction in predicting customer churn in B2B,” *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00500-3.
- [30] A. F. Cunha, D. Ferreira, C. Neto, A. Abelha, and J. Machado, “A CRISP-DM Approach for Predicting Liver Failure Cases: An Indian Case Study,” *Lect. Notes Networks Syst.*, vol. 271, pp. 156–164, 2021, doi: 10.1007/978-3-030-80624-8_20.
- [31] C. J. Costa and J. T. Aparicio, “A Methodology to Boost Data Science in the Context of COVID-19,” *Adv. Parallel Distrib. Process. ...*, 2021, doi: https://doi.org/10.1007/978-3-030-69984-0_7.
- [32] D. Kurniawan and M. Yasir, “Optimization Sentimen Analysis using CRISP-DM and Naive Bayes Methods Implemented on Social Media,” *Cybersp. J. Pendidik. Teknol. Inf.*, vol. 6, no. 2, p. 74, 2022, doi: 10.22373/cj.v6i2.12793.
- [33] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, “Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM,” *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, Mar. 2022, doi: 10.34010/jati.v12i1.6674.
- [34] Y. A. Singgalen, “Analisis Performa Algoritma NBC, DT, SVM dalam Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2766.
- [35] S. Shedriko and M. Firdaus, “Penentuan Klasifikasi Dengan Crisp-Dm Dalam Memprediksi Kelulusan Mahasiswa Pada Suatu Mata Kuliah,” *Semnas Ristek (Seminar Nas. Ris. dan Inov. Teknol.)*, vol. 6, no. 1, 2022, doi: 10.30998/semnasristek.v6i1.5814.
- [36] T. Wuriyanto, H. B. Setiawan, and A. B. Tjandrarini, “Penerapan Model CRISP-DM pada Prediksi Nasabah Kredit yang Berisiko Menggunakan Algoritma Support Vector Machine,” *J. Ilm. Scroll (Jendela Teknol. Informasi)*, vol. 10, no. 1, pp. 1–6, 2022, [Online]. Available: <https://univ45sby.ac.id/ejournal/index.php/informatika/article/view/291>
- [37] A. K. Rotty, “Cross-Industry Standard Process

- for Data Mining (CRISP-DM) Approach in Determining the Most Significant Employee Engagement Drivers to Sales at X Car Dealership,” pp. 3368–3379, 2023, doi: 10.46254/ap03.20220552.
- [38] M. Rafi Muttaqin, T. Iman Hermanto, M. Agus Sunandar, P. Studi Teknik Informatika, and S. Tinggi Teknologi Wastukencana, “Penerapan K-Means Clustering dan Cross-Industry Standard Process For Data Mining (CRISP-DM) untuk Mengelompokan Penjualan Kue,” *Journal.Unpak.Ac.Id*, vol. 191, Rafi, no. 1, pp. 38–53, 2022, [Online]. Available: <http://journal.unpak.ac.id/index.php/komputasi/article/view/3976>
- [39] V. Plotnikova, M. Dumas, and F. P. Milani, “Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements,” *Data Knowl. Eng.*, vol. 139, 2022, doi: 10.1016/j.datak.2022.102013.
- [40] J. A. Rocha *et al.*, “Human Activity Recognition Through Wireless Body Sensor Networks (WBSN) Applying Data Mining Techniques,” *Smart Innov. Syst. Technol.*, vol. 253, pp. 327–339, 2022, doi: 10.1007/978-981-16-5036-9_31.
- [41] L. Drowatzky, H. Wiemer, and S. Ihlenfeldt, “Data Mining Suitable Digitization of Production Systems – A Methodological Extension to the DMME,” *Lect. Notes Prod. Eng.*, vol. Part F1163, pp. 524–534, 2023, doi: 10.1007/978-3-031-18318-8_53.
- [42] F. Alzyoud, “Improved model for traffic management system using KDD and big data: case study Jordan traffic,” *Int. J. Comput. Commun. Control*, vol. 18, no. 3, May 2023, doi: 10.15837/ijccc.2023.3.5006.
- [43] V. H. Ortega-Guzmán, L. Gutiérrez-Preciado, F. Cervantes, and M. Alcaraz-Mejia, “A Methodology for Knowledge Discovery in Labeled and Heterogeneous Graphs,” *Appl. Sci.*, vol. 14, no. 2, p. 838, Jan. 2024, doi: 10.3390/app14020838.
- [44] A. Pambudi, “Penerapan Crisp-Dm Menggunakan Mlr K-Fold Pada Data Saham Pt. Telkom Indonesia (Persero) Tbk (Tlkm) (Studi Kasus: Bursa Efek Indonesia Tahun 2015-2022),” *J. Data Min. dan Sist. Inf.*, vol. 4, no. 1, p. 1, 2023, doi: 10.33365/jdmsi.v4i1.2462.
- [45] Y. A. Singgalen, “Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT),” *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 3, p. 1551, Jul. 2023, doi: 10.30865/mib.v7i3.6461.
- [46] Y. A. Singgalen, “Analisis Sentimen Pengunjung Pulau Komodo dan Pulau Rinca di Website Tripadvisor Berbasis CRISP-DM,” *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 614–625, Jan. 2023, doi: 10.47065/josh.v4i2.2999.
- [47] L. Rahmadi, Hadiyanto, R. Sanjaya, and A. Prambayun, “Crop Prediction Using Machine Learning with CRISP-DM Approach,” 2023, pp. 399–421. doi: 10.1007/978-981-99-6550-2_31.
- [48] D. U. Iswavigra, L. E. Zen, and H. Hanim, “Marketing Strategy UMKM Dengan CRISP-DM Clustering &Promotion Mix Menggunakan Metode K-Medoids,” *J. Inf. dan Teknol.*, vol. 5, pp. 45–54, 2023, doi: <https://doi.org/10.37034/jidt.v5i1.260>.
- [49] F. Lee, J. M. Robinson, and C. Baxter, “Using Supervised Machine Learning and CRISP-DM to Predict an Acquittal Verdict,” *Inf. Syst. Educ. J.*, vol. 21, no. 4, pp. 23–36, 2023, [Online]. Available: <https://isedj.org/>; <https://iscap.info>
- [50] N. C. Sastya and I. Nugraha, “Penerapan Metode CRISP-DM dalam Menganalisis Data untuk Menentukan Customer Behavior di MeatSolution,” *Unistek*, vol. 10, no. 2, pp. 103–115, 2023, doi: 10.33592/unistek.v10i2.3079.
- [51] Y. A. Singgalen, “Penerapan Metode CRISP-DM untuk Optimalisasi Strategi Pemasaran STP (Segmenting, Targeting, Positioning) Layanan Akomodasi Hotel, Homestay, dan Resort,” *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 4, p. 1980, Oct. 2023, doi: 10.30865/mib.v7i4.6896.
- [52] U. Gupta and R. Sharma, “Analysis of criminal spatial events in india using exploratory data analysis and regression,” *Comput. Electr. Eng.*, vol. 109, p. 108761, Jul. 2023, doi: 10.1016/j.compeleceng.2023.108761.
- [53] B. Mirza *et al.*, “A clinical site workload prediction model with machine learning lifecycle,” *Healthc. Anal.*, vol. 3, p. 100159, Nov. 2023, doi: 10.1016/j.health.2023.100159.
- [54] J. Weller, D. Roesmann, S. Eggert, S. von Enzberg, I. Gräßler, and R. Dumitrescu, “Identification and prediction of standard times in machining for precision steel tubes through the usage of data analytics,” *Procedia CIRP*, vol. 119, pp. 514–520, 2023, doi: 10.1016/j.procir.2023.01.011.
- [55] YA Singgalen, “Utilizing Knowledge Discovery in Databases (KDD) for Hotel Guest Feedback Analysis,” 2024, doi: <http://dx.doi.org/10.47065/josyc.v6i1.6094>.
- [56] Y. A. Singgalen, “Understanding Hotel Customer Experience through User-Generated Reviews using Knowledge Discovery in Databases (KDD) Understanding Hotel Customer Experience through User-Generated Reviews using Knowledge Discovery in Databases (KDD),” no. November, 2024, doi: 10.47065/josyc.v6i1.6014.
- [57] C. Llatas, B. Soust-Verdaguer, L. C. Torres, and D. Cagigas, “Application of Knowledge Discovery in Databases (KDD) to environmental, economic, and social indicators used in BIM workflow to support sustainable design,” *J. Build. Eng.*, vol. 91, p. 109546, Aug. 2024, doi:

- 10.1016/j.jobe.2024.109546.
- [58] J. Bokrantz, M. Subramaniyan, and A. Skoogh, "Realising the promises of artificial intelligence in manufacturing by enhancing CRISP-DM," *Prod. Plan. Control*, vol. 35, no. 16, pp. 2234–2254, 2023, doi: 10.1080/09537287.2023.2234882.
- [59] A. A. RH Bemthuis, RR Govers, "A CRISP-DM-based Methodology for Assessing Agent-based Simulation Models using Process Mining," 2024, doi: <https://doi.org/10.48550/arXiv.2404.01114>.
- [60] Y. A. Singgalen, "Sentiment Classification of The Capsule Hotel Guest Reviews using Cross-Industry Standard Process for Data Mining (CRISP-DM)," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 1, p. 632, Jan. 2024, doi: 10.30865/mib.v8i1.7329.
- [61] Y. A. Singgalen, "Implementation of CRISP-DM for Social Network Analysis (SNA) of Tourism and Travel Vlog Content Reviews," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 1, p. 572, Jan. 2024, doi: 10.30865/mib.v8i1.7323.
- [62] P. Fernandes and M. C. Coelho, "Data mining techniques for estimating instantaneous NOX emissions from Euro 6 diesel passenger cars," *Transp. Res. Procedia*, vol. 78, pp. 444–451, 2024, doi: 10.1016/j.trpro.2024.02.056.
- [63] R. Sousa, M. Passos, M. Almeida, M. Ribeiro, and H. Peixoto, "Harnessing Data Mining to Predict Survival Outcomes in Patients with Hepatic Cirrhosis," *Procedia Comput. Sci.*, vol. 238, pp. 938–943, 2024, doi: 10.1016/j.procs.2024.06.116.
- [64] L. Z. S. Sudar, J. L. Imbenay, I. Budi, A. Ramadiah, P. K. Putra, and A. B. Santoso, "Textual Analysis for Public Sentiment Toward National Police Using CRISP-DM Framework," *Rev. d'Intelligence Artif.*, vol. 38, no. 1, pp. 63–72, Feb. 2024, doi: 10.18280/ria.380107.
- [65] S. Mirfakhraei, N. Abdolvand, and S. R. Harandi, "The RFMRv Model for Customer Segmentation Based on the Referral Value," *Interdiscip. J. Manag. Stud.*, vol. 17, no. 2, pp. 455–473, 2024, doi: 10.22059/ijms.2023.329229.674722.
- [66] M. Karmagatri, C. I. Ratnapuri, and B. Casteillo, "Identifying Key Behaviour Patterns that Influence Plastic Bag Refusal Decisions," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1324, no. 1, p. 012081, Apr. 2024, doi: 10.1088/1755-1315/1324/1/012081.
- [67] Y. Wu, D. Zhao, J. Peng, X. Xiang, and H. Huang, "Hybrid Electric Vehicle Powertrain Mounting System Optimization Based on Cross-Industry Standard Process for Data Mining," *Electronics*, vol. 13, no. 6, p. 1117, Mar. 2024, doi: 10.3390/electronics13061117.
- [68] E. Muningsih, F. F. D. Imaniawan, A. Widayanto, E. A. Pratama, S. Sutrisno, and S. Kiswati, "Optimized decision tree classification method for diabetes prediction," *TELKOMNIKA (Telecommunication Comput. Electron. Control)*, vol. 22, no. 4, p. 941, Aug. 2024, doi: 10.12928/telkomnika.v22i4.25656.