IT-EXPLORE
Jurnal Penerapan Teknologi Informasi dan Komunikasi
ISSN 2828-7940, e-ISSN 2829-1727, DOI: 10.24246/itexplore.v4i2.2025.pp138-149
Volume 04 Nomor 02 Tahun 2025
Journal homepage: https://ejournal.uksw.edu/itexplore

Teknik dan aplikasi data mining di Indonesia: tinjauan literatur satu dekade (2015-2024)

Eliana Saputri[1]
[1] Fakultas Sains dan Teknologi, Universitas Islam negeri Syarif Hidayatullah Jakarta

Informasi Artikel

Riwayat Artikel:

Dikirim:
18 Januari 2025
Direvisi:
25 Februari 2025
Diterima:
26 Juni 2025

Kata Kunci:

Data Mining,
Industry Applications of Data Mining,
PICO,
Systematic Literature Review

Keywords:

Data Mining,
Industry Applications of Data Mining,
PICO,
Systematic Literature Review

This is an open access article under the CC BY-SA license

ABSTRAK

Pentingnya *data mining* di Indonesia semakin meningkat seiring dengan pertumbuhan data besar di berbagai sektor strategis. *Data mining* memainkan peran penting dalam memproses data yang kompleks menjadi suatu informasi yang bermanfaat digunakan untuk mendukung pengambilan keputusan berbasis data, yang sangat dibutuhkan dalam menghadapi tantangan kompetitif dan kompleksitas operasional. Penelitian ini bertujuan untuk mengkaji perkembangan teknik dan aplikasi data mining di Indonesia selama dekade terakhir (2015-2024). Melalui pendekatan *systematic literature review*, penelitian ini menganalisis tren penggunaan berbagai teknik data mining di sejumlah sektor industri. Data dikumpulkan dari publikasi akademik dalam database terindeks SCOPUS untuk mengidentifikasi teknik yang paling umum digunakan serta aplikasinya dalam setiap sektor. Dari 95 paper awal yang ditemukan, diseleksi kembali berdasarkan ketersediaan akses, judul, dan abstrak, hingga didapatlah 64 paper. Hasil penelitian menunjukkan bahwa teknik seperti *K-Means, Naive Bayes*, dan *Decision Tree* paling sering digunakan. Di sektor bisnis, clustering melalui K-Means banyak diterapkan untuk segmentasi pasar dan analisis pola konsumen. Sektor kesehatan terutama memanfaatkan teknik *classification*, seperti Naive Bayes dan Decision Tree, untuk prediksi risiko penyakit dan diagnosis awal. Sementara itu, sektor pendidikan menggunakan data mining untuk menilai kinerja siswa dan memprediksi potensi *dropout*, membantu institusi dalam mengoptimalkan strategi pembelajaran.

ABSTRACT

The importance of data mining in Indonesia is increasing along with the growth of big data in various strategic sectors. Data mining plays an important role in transforming complex data into useful information to support data-driven decision making, which is urgently needed in the face of competitive challenges and operational complexity. This research aims to examine the development of data mining techniques and applications in Indonesia over the last decade (2015-2024). Through a systematic literature review approach, data was collected from academic publications in SCOPUS indexed databases. From the initial 95 papers found, a further selection was made based on accessibility, title, and abstract until 64 papers were included in the article review. The results show that techniques such as K-Means, Naive Bayes, and Decision Tree are most commonly used. In the business sector, clustering through K-Means is widely applied for market segmentation and consumer pattern analysis. The healthcare sector mainly utilizes classification techniques, such as Naive Bayes and Decision Tree, for disease risk prediction and early diagnosis. Meanwhile, the education sector uses data mining to assess student performance and predict potential dropouts, assisting institutions in optimizing learning strategies.

Corresponding Author:
Eliana Saputri
Fakultas Sains dan Teknologi, Universitas Islam negeri Syarif Hidayatullah Jakarta
Jl. Ir. H. Djuanda No. 95, Ciputat Timur, Tangerang Selatan 15412, Banten
Email : eliana.saputri22@mhs.uinjkt.ac.id

1. PENDAHULUAN

Seiring dengan meningkatnya *volume* dan variasi data, *data mining* telah berkembang secara signifikan selama satu dekade terakhir (2015-2024). Hal ini didorong oleh meningkatnya penggunaan teknologi digital, *Internet of Things* (*IoT*), media sosial, dan sistem berbasis *cloud*. Dengan semakin kompleksnya data yang tersedia, metode analisis konvensional menjadi kurang efektif, sehingga diperlukan teknik yang lebih canggih untuk mengekstrak informasi yang bermakna. *Data mining* hadir sebagai salah satu pendekatan utama dalam mengolah data besar (*big data*) untuk mendapatkan pola, hubungan, dan wawasan yang dapat digunakan dalam pengambilan keputusan berbasis data [1]. *Data mining*, atau penambangan data, menjadi salah satu teknologi yang berkembang pesat dan memungkinkan pengguna untuk menggali informasi berharga dari kumpulan data yang kompleks. *Data mining* tidak hanya bermanfaat dalam mengolah data secara efisien tetapi juga mampu mendukung pengambilan keputusan berbasis data, yang kini menjadi dasar dalam berbagai bidang, seperti bisnis, kesehatan, keuangan, dan pendidikan.

Data mining, atau penambangan data, merupakan teknik yang memanfaatkan berbagai metode statistika, pembelajaran mesin (*machine learning*), dan kecerdasan buatan (*artificial intelligence*) untuk mengidentifikasi pola tersembunyi dalam data. Penerapan data mining telah terbukti efektif dalam berbagai bidang, seperti bisnis, kesehatan, keuangan, pendidikan, dan pemerintahan. Seperti teknik *clustering* yang terus berkembang dan menjadi salah satu metode penting dalam analisis data, terutama dalam sektor keuangan dan pemasaran [2]. Selanjutnya teknik klasifikasi mengkategorikan data berdasarkan karakteristik yang telah ditentukan, menggunakan metode seperti *decision tree*, *random forest*, dan *support vector machine* [3]. Lalu ada juga *Association Rule Mining* untuk menemukan pola asosiasi antara variabel, misalnya dalam analisis pembelian konsumen untuk menentukan barang yang sering dibeli bersama [4]. Teknik *forecasting* memprediksi nilai atau tren masa depan berdasarkan pola historis data. Contohnya adalah penggunaan metode ARIMA dan *exponential smoothing* dalam analisis *time series* untuk prediksi penjualan atau peramalan cuaca [1]. Dan terakhir teknik estimasi memungkinkan penentuan nilai tertentu berdasarkan pola data yang ada. Estimasi sering digunakan dalam peramalan harga, analisis penjualan, dan penentuan *rating*. Algoritma regresi, seperti regresi linier, sering diterapkan untuk tujuan ini [4].

Meskipun berbagai penelitian telah dilakukan, belum banyak kajian sistematis yang secara komprehensif memetakan penerapan teknik *data mining* di Indonesia. Sebagian besar studi cenderung berfokus pada aplikasi tunggal di sektor tertentu tanpa menyoroti perkembangan teknologi ini secara holistik. Oleh karena itu, tujuan dari kajian ini adalah untuk mengidentifikasi berbagai teknik *data mining* yang telah diterapkan dalam berbagai sektor di Indonesia dalam satu dekade terakhir, menganalisis tren perkembangan dan penerapan data mining di berbagai industri, serta menyajikan pemetaan penelitian terdahulu terkait teknik dan aplikasi *data mining* di Indonesia. Selain itu, penelitian ini juga diharapkan dapat memberikan wawasan bagi akademisi, praktisi industri, serta pemangku kepentingan mengenai perkembangan dan manfaat *data mining* dalam pengolahan data besar di Indonesia. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang *data mining* serta menjadi referensi bagi penelitian selanjutnya dalam mengeksplorasi teknik dan aplikasi *data mining* di Indonesia. Dengan adanya kajian ini, diharapkan dapat memberikan landasan bagi kebijakan berbasis data yang lebih baik dan mendukung perkembangan industri berbasis teknologi di Indonesia.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Sejumlah penelitian dalam dekade terakhir telah membandingkan berbagai teknik data mining untuk mengeksplorasi efektivitasnya dalam berbagai aplikasi di Indonesia. Penelitian-penelitian ini penting untuk memahami kelebihan dan kekurangan setiap teknik data mining dalam menghadapi tantangan dan

kompleksitas data yang berbeda.

Sugiyarto membandingkan tiga algoritma klasifikasi, yaitu D*ecision Tree*, *Naive Bayes*, dan *Support Vector Machine* (SVM), dalam konteks prediksi risiko kredit di Indonesia [5]. Hasil penelitian ini menunjukkan bahwa *SVM* memiliki akurasi yang lebih tinggi dibandingkan *decision tree* dan *naive Bayes*, meskipun dengan waktu pelatihan yang lebih lama. Penelitian ini memberikan wawasan tentang pentingnya memilih algoritma yang tepat berdasarkan trade-off antara akurasi dan kecepatan pemrosesan, serta kompleksitas data yang terlibat.

Di sisi lain, Yuda melakukan perbandingan antara teknik *k-means*, *DBSCAN*, dan *hierarchical clustering* dalam aplikasi segmentasi pasar di Indonesia. Mereka menemukan bahwa meskipun *k-means* lebih cepat dalam memproses data yang lebih besar, *DBSCAN* lebih efektif dalam menangani data dengan bentuk cluster yang kompleks dan tidak teratur. *Hierarchical clustering*, sementara itu, unggul dalam mengidentifikasi hubungan hierarkis dalam data, meskipun membutuhkan waktu pemrosesan yang lebih lama. Penelitian ini menekankan bahwa pemilihan teknik clustering yang tepat sangat bergantung pada struktur data dan tujuan analisis [6].

Selain itu, Madhuri membandingkan teknik regresi dan klasifikasi dalam memprediksi harga properti di Indonesia. Dalam penelitian ini, mereka menguji *linear regression*, *decision tree regression*, dan *random forest* untuk menilai ketepatan prediksi harga properti. Hasilnya menunjukkan bahwa *random forest* menghasilkan prediksi yang lebih akurat dibandingkan *linear regression* dan *decision tree regression*, meskipun model *random forest* lebih kompleks. Penelitian ini juga menggambarkan pentingnya memilih teknik yang sesuai dengan karakteristik dan kompleksitas data yang digunakan [7].

Widyawati dalam penelitiannya melakukan studi perbandingan pada aplikasi prediksi kanker menggunakan teknik klasifikasi seperti *k-nearest neighbor* (k-NN), *naive Bayes*, dan *random forest*. Penelitian ini berfokus pada dataset medis yang mencakup informasi pasien di Indonesia dan menunjukkan bahwa *random forest* memberikan akurasi yang lebih tinggi dalam memprediksi jenis kanker, dibandingkan dengan k-NN dan *naive Bayes*. Hasil ini menunjukkan bahwa meskipun *random forest* lebih kompleks, ia memberikan kinerja yang lebih baik dalam kasus-kasus yang melibatkan dataset besar dan variabel yang banyak [8].

Dalam konteks prediksi lalu lintas di kota-kota besar Indonesia, Tamir membandingkan penggunaan algoritma regresi linier dan *artificial neural networks* (ANN) untuk memprediksi kondisi lalu lintas berdasarkan data dari sensor lalu lintas dan aplikasi GPS. Hasil penelitian menunjukkan bahwa *ANN* memiliki kemampuan yang lebih baik dalam menangani data non-linier dan memberikan prediksi yang lebih akurat dibandingkan dengan regresi linier, meskipun memerlukan waktu pelatihan yang lebih lama. Penelitian ini menyoroti keunggulan teknik *ANN* dalam aplikasi yang memerlukan pemodelan data yang lebih kompleks [9].

Mariel dalam penelitiannya membandingkan efektivitas teknik *deep learning* dan algoritma *support vector machine* (SVM) dalam analisis sentimen terhadap ulasan produk di Indonesia. Mereka menemukan bahwa *deep learning* memberikan akurasi yang lebih tinggi dalam menangkap nuansa emosi dan opini pengguna dibandingkan dengan SVM, yang cenderung lebih terbatas dalam menangani data teks yang kompleks. Penelitian ini menunjukkan potensi besar penggunaan *deep learning* dalam aplikasi analisis teks dan sentimen, yang semakin berkembang di Indonesia [10].

Secara keseluruhan, penelitian-penelitian ini menunjukkan bahwa tidak ada satu teknik data mining yang dapat dianggap unggul secara mutlak. Pemilihan teknik yang tepat sangat bergantung pada karakteristik data, tujuan analisis, dan sumber daya yang tersedia. Sebagai hasilnya, perbandingan teknik-teknik data mining terus menjadi bagian integral dalam penelitian dan penerapan teknologi data mining di Indonesia, dengan harapan dapat memberikan solusi yang lebih efisien dan akurat di berbagai bidang.

2.2. Landasan Teori

Data mining, atau penambangan data, adalah proses analisis data besar-besaran untuk menemukan pola, hubungan, atau pengetahuan baru yang berguna bagi pengambilan keputusan. Data mining bukan sekadar mengekstraksi data, tetapi juga melibatkan pemahaman, pembersihan, dan transformasi data agar hasil

analitik dapat bermanfaat dalam konteks yang lebih luas. Proses ini sering kali menggunakan algoritma statistik, *machine learning*, dan kecerdasan buatan untuk mengidentifikasi informasi tersembunyi di dalam data yang kompleks.

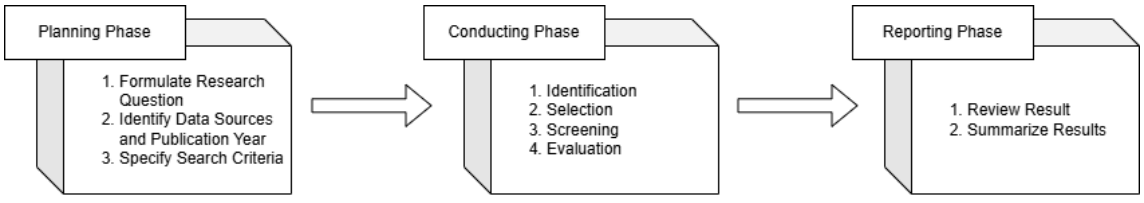
Menurut Witten, *data mining* terdiri dari beberapa komponen kunci, yaitu ekstraksi data, transformasi data, dan analisis pola. Ekstraksi data melibatkan pengumpulan dan penggabungan data dari berbagai sumber [1]. Transformasi data kemudian mengubah data mentah menjadi format yang sesuai untuk dianalisis, seperti normalisasi dan penghapusan noise. Analisis pola, yang menjadi inti *data mining*, mencakup penerapan algoritma untuk menemukan pola atau hubungan dalam data.

Ada beberapa teknik utama dalam data mining yang sering digunakan dalam berbagai aplikasi, antara lain:

1. *Classification*: Teknik ini mengkategorikan data berdasarkan karakteristik yang telah ditentukan, menggunakan metode seperti *decision tree*, *random forest*, dan *support vector machine* [3].
2. *Clustering*: Teknik *clustering* mengelompokkan data ke dalam grup berdasarkan kemiripan, misalnya dengan algoritma k-means atau DBSCAN. Sazonov mencatat bahwa clustering bermanfaat dalam segmentasi pasar dan analisis jaringan social [2].
3. *Association Rule Mining*: Teknik ini menemukan pola asosiasi antara variabel, misalnya dalam analisis pembelian konsumen untuk menentukan barang yang sering dibeli bersama [4].
4. *Forecasting*: Teknik *forecasting* memprediksi nilai atau tren masa depan berdasarkan pola historis data. Contohnya adalah penggunaan metode ARIMA dan *exponential smoothing* dalam analisis *time series* untuk prediksi penjualan atau peramalan cuaca [1].
5. *Estimation*: Teknik estimasi memungkinkan penentuan nilai tertentu berdasarkan pola data yang ada. Estimasi sering digunakan dalam peramalan harga, analisis penjualan, dan penentuan rating. Algoritma regresi, seperti regresi linier, sering diterapkan untuk tujuan ini [4].

3. METODE PENELITIAN

Penelitian ini menggunakan metode *systematic review*. Metode *systematic review* ini sendiri merupakan proses primer yang mendasari pembuatan artikel, dan akan berdampak besar bagi hasil penelitian. Menurut artikel mengenai *systematic literature review*, dikatakan bahwa ada 3 jenis *literature review* [11]. Pertama, adalah *theoretical background* atau bisa disebut review atas landasan konseptual dan teoritis bagi penelitian yang akan dilakukan. Kedua, ada *thesis literature review* atau bisa disebut peninjauan landasan konseptual dan teoritis bagi penelitian yang akan dilakukan. Lalu yang terakhir adalah *standalone literature review* yakni peninjauan literatur mengenai topik tertentu, tanpa menggabungkannya dengan penelitian empiris atau eksperimen yang terbit pada waktu yang baru. Adapun penelitian ini akan menggunakan jenis yang kedua yakni *thesis literature review*.



Gambar 1. Tahapan dalam SLR

Systematic literature review dilaksanakan secara bertahap. Adapun tahapan dalam SLR ini adalah *planning* (perencanaan), *conducting* (pelaksanaan), dan juga *reporting* (pelaporan) [12]. Tahap pertama yakni tahap perencanaan, yang pada penelitian kali ini sudah disertakan pada bagian pendahuluan, yang termasuk didalamnya latar belakang penelitian, pembahasan mengenai *data mining*, dan juga tujuan penelitian. Selanjutnya adalah tahap pelaksanaan yang akan dilakukan dengan cara mencari, menyaring, menilai kualitas, lalu menyatukan data-data mengenai penggunaan data mining di Indonesia yang telah didapat. Terakhir, tentunya suatu SLR tidak akan rampung tanpa pelaporan atau bisa juga disebut pelaporan atas tinjauan pustaka yang akan disajikan dalam bagian hasil dan pembahasan.

3.1. Pertanyaan Penelitian

Pertanyaan penelitian ini dibuat dengan bantuan struktur PICO yang terdiri dari *Population* (Populasi), *Intervention* (Intervensi), *Comparison* (Perbandingan), dan *Outcome* (Hasil). PICO digunakan untuk merumuskan *research question* dan *keyword* pencarian studi tentang teknik *data mining* dan aplikasinya. Tabel 1 di bawah ini menjelaskan struktur PICO yang penulis gunakan.

Tabel 1 *Structure of PICO*

PICO	*Details*
Population	Studi *data mining* di Indonesia, 2015-2024
Intervention	Teknik *data mining* dan aplikasinya
Comparison	Tidak ada
Outcome	Tren dan wawasan *data mining* untuk pengembangan lokal

Selanjutnya, menentukan *research question* dan tujuan dari setiap pertanyaan penelitian, seperti yang terlampir pada Tabel 2.

Tabel 2 *Research Question*

ID	*Research Question*	*Purpose*
RQ1	Apa saja teknik *data mining* utama yang telah diterapkan di Indonesia selama satu dekade terakhir (2015-2024)?	Mengidentifikasi teknik data mining utama yang telah diterapkan di Indonesia dalam rentang waktu 2015-2024.
RQ2	Apa tren utama dalam aplikasi data mining di sektor-sektor tertentu di Indonesia (misalnya, kesehatan, bisnis, pendidikan, dll.) dari 2015 hingga 2024?	Menjelaskan tren dan perkembangan aplikasi data mining di sektor-sektor spesifik di Indonesia, seperti kesehatan, bisnis, dan pendidikan, dll.

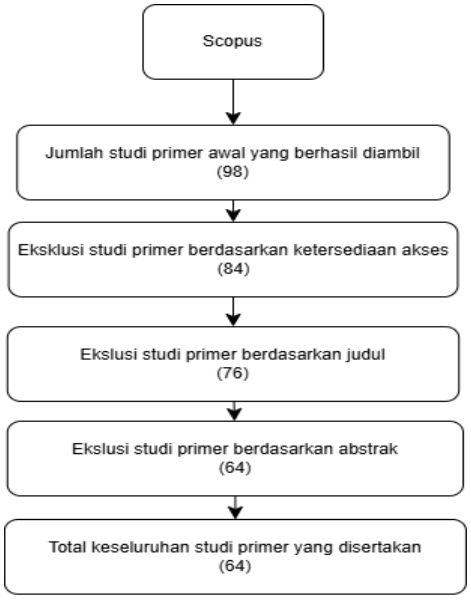
B. Strategi pencarian

Strategi pencarian yang terdapat pada penelitian mengenai data mining ini, menggunakan kata kunci dalam melakukan proses pencariannya. Pencarian dilakukan dengan menggunakan bantuan dari struktur PICO yang telah dibuat sebelumnya, khususnya dalam mengidentifikasi istilah ada pada bagian *Population* dan *Intervention*, langkah berikutnya yakni mengidentifikasi istilah berdasarkan pada pertanyaan-pertanyaan yang ada dalam penelitian, kemudian meneliti pertanyaan yang telah ditentukan dan mengidentifikasi kata-kata sinonim untuk istilah pencarian.

Berikut ini adalah kata kunci yang telah berhasil ditentukan dengan menggabungkan setiap istilah yang diambil dengan menggunakan operator logika supaya bisa didapatkan hasil pencarian yang relevan yaitu "*Data Mining Technique*" *OR* "*Data Mining Application*" *AND* "*Indonesia*". Berdasarkan kata kunci yang telah ditentukan, dilakukannya pencarian dengan bantuan Publish or Perish dengan *database* ilmiah digital SCOPUS. Dari pencarian pada *database* tersebut, pencarian terbatas pada jurnal artikel dan konferensi di Indonesia serta diterbitkan pada tahun 2015 sampai dengan 2024.

C. Strategi Seleksi Kriteria Inklusi dan Eksklusi

Dalam proses seleksi studi awal telah ditetapkan kriteria inklusi yaitu seluruh *journal article* dan konferensi, serta publikasi penelitian yang berhubungan dengan *data mining* di Indonesia. Kriteria eksklusi yang ditetapkan adalah publikasi non-ilmiah dan publikasi penelitian mengenai data mining di negara lain dan di luar lingkup penelitian ini. Hasil total pencarian studi primer awal yang berhasil diambil adalah 95 *paper*, kemudian diseleksi kembali berdasarkan ketersediaan akses menjadi 84 *paper*, selanjutnya diseleksi kembali berdasarkan judul menjadi 76 *paper*, dan terakhir diseleksi berdasarkan abstrak, didapatlah 64 *paper* yang disertakan dalam review artikel.



Gambar 2. Hasil total keseluruhan studi primer yang disertakan dalam penelitian

4. HASIL DAN PEMBAHASAN

Berdasarkan subbab yang sudah dijelaskan sebelumnya, terdapat sejumlah 64 *paper* dipilih dari 95 *paper awal* menggunakan metode *systematic literature review*. *Paper* yang dipilih tersebut dirangkum dalam Tabel 3 berikut ini.

Tabel 3 *Selected Papers*

Author	Judul
[13]	*"Fishing gear recognition from VMS data to identify illegal fishing activities in Indonesia"*
[14]	*"Predicting students achievement based on motivation in vocational school using data mining approach"*
[15]	*"Application of Decision Tree Approach to Student Selection Model- A Case Study"*
[16]	*"Comparative analysis of classification methods in determining non-active student characteristics in Indonesia Open University"*
[17]	*"The clustering of high schools based on national and school examinations: A case study at Daerah Istimewa Yogyakarta Province"*
[18]	*"Identifying consumer buying behavior differences through market basket analysis in multiple outlet types"*
[19]	*"The Application of Data Mining Techniques to Create Promotion Strategy for Mobile Phone Shop"*
[20]	*"Data mining application to detect financial fraud in Indonesia's public companies"*
[21]	*"Comparative study on data mining classification methods for cervical cancer prediction using pap smear results"*
[22]	*"Predictive data mining of chronic diseases using decision tree: A case study of health insurance company in Indonesia"*
[23]	*"Analysis of earthquake magnitude level based on data Twitter with decision tree algorithm"*
[24]	*"An analysis of natural disaster data by using K-means and K-medoids algorithm of data mining techniques"*
[25]	*"A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques"*
[26]	*"ST-DBSCAN clustering module in SpagoBI for hotspots distribution in Indonesia"*
[27]	*"Integration K-Means Clustering Method and Elbow Method for Identification of the*

Best Customer Profile Cluster"
[28] *"Data mining technique with cluster anaysis use K-means algorithm for LQ45 index on Indonesia stock exchange"*
[29] *"Maximum attribute relative approach of soft set theory in selecting cluster attribute of electronic government data set"*
[30] *"Review on fuzzy expert system and data mining techniques for the diagnosis of coronary artery disease"*
[31] *"A Relative Tolerance Relation of Rough Set (RTRS) for Potential Fish Yields in Indonesia"*
[32] *"Detection of interviewer falsification in statistics Indonesia's mobile survey*
[33] *"Forecasting the welfare of fishermen and aquaculture farmers in Indonesia: Data Mining Approach"*
[34] *"Automatic rule generator via FP-growth for eye diseases diagnosis"*
[35] *"Developing an Indonesia's health literacy short-form survey questionnaire (HLS-EU-SQ10-IDN) using the feature selection and genetic algorithm"*
[36] *"Analysis and prediction of diabetes complication disease using data mining algorithm"*
[37] *"Implementation of Data Mining for Determining Majors Using K-Means Algorithm in Students of SMA Negeri 1 Pangkalan Kerinci"*
[38] *"Predicting click-through rates using data mining technique on digital advertisements"*
[39] *"Promo abuse modeling in e-commerce using machine learning approach"*
[40] *"Application of Data Mining Technique using K-Medoids in the case of Export of Crude Petroleum Materials to the Destination Country"*
[41] *"Customer churn prediction models for PT. XYZ Insurance"*
[42] *"Implementation of Data Mining in Grouping Percentage of Blind Letters Age 15+ by Province Using K-Means Algorithm"*
[43] *"Data mining classification approach to predict the duration of contraceptive use"*
[44] *"Association Rule Implementation Using Algorithm Apriori to Analize Fishing Pattern in Indonesia"*
[45] *"Performance Comparison of Data Mining Techniques for Rain Prediction Models in Indonesia"*
[46] *"Predictability of Naïve Bayes classifier for lahar hazard mapping by weather radar"*
[47] *"Utilization of Data Mining Techniques in National Food Security during the Covid-19 Pandemic in Indonesia"*
[48] *"Implementation of data mining with Apriori techniques to determine the pattern of purchasing of agricultural equipment (Case Study: XYZ Store)"*
[49] *"Application of the clustering algorithm to the small and micro industrial companies for mapping regions with k-medoids"*
[50] *"Heart Disease Classification Model Using K-Nearest Neighbor Algorithm"*
[51] *"Application of data mining techniques using the K-Means Method on Unmet Need of Health Services by Province in Indonesia"*
[52] *"Utilization of the field of data mining in mapping the area of the Human Development Index (HDI) in Indonesia"*
[53] *"Formulation of city health development index using data mining"*
[54] *"Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms"*
[55] *"Analysis of the effect of the lecturer satisfaction with the Naive Bayes Data Mining technique on institutional performance"*
[56] *"Predicting Student's Failure in Education Based on Dropout Status"*
[57] *"Cluster Application with K-Means Algorithm on the Population of Trade and Accommodation Facilities in Indonesia"*
[58] *"Data mining techniques for predicting the financial performance of Islamic banking in Indonesia"*
[59] *"Predictive modelling of mercury pollution in Indonesian ASGM areas and evaluation of the main factors using machine learning algorithm"*
[60] *"Prediction of Mortality in Trauma Patients with Insufficient Training Data Using Deep Learning"*

[61] *"Game innovation: a case study using the Kizzugemu visual novel game with Tyranobuilder software in elementary school"*
[62] *"Application of Data Mining for Slot Time Prediction at International Airports in Indonesia: J48 Algorithm"*
[63] *"State-of-Art on Data Mining and Microscopic Simulation Techniques for Evaluation of Urban Uncontrolled Intersections"*
[64] *"Implementation Frequent Pattern Growth (FP-Growth) Algorithm on Market Basket Analysis for Determining Consumer Purchasing Patterns (Case Study: XYZ Coffee Shop)"*
[65] *"Predicting Stock Market Trends with News Sentiment Analysis and Historical Stock Market Data Analysis Using Data Mining Techniques"*
[66] *"Data mining framework for the identification of profitable customer based on recency, frequency, monetary (RFM)"*
[67] *"Analysis K-Means of Covid-19 Cases in Bandar Lampung and South Lampung"*
[68] *"Comparison of Decision Tree and Support Vector Machine for Predicting Jakarta Air Quality Index"*
[69] *"Classification of Food Distribution Status and Pattern Mapping of Optimal Distribution in West Java"*
[70] *"K-Nearest Neighbors Method Prediction of the Anti-Corruption Behavior Index by Region of Residence"*
[71] *"How can we predict transportation stock prices using artificial intelligence? Findings from experiments with Long Short-Term Memory based algorithms"*
[72] *"Predicting daily precision improvement of Jakarta Islamic Index in Indonesia's Islamic stock market using big data mining"*
[73] *"Classification of Salt Quality Based on the Content of Several Elements in the Salt Using Machine Learning"*
[74] *"Time Series Clustering Analysis for Increases Food Commodity Prices in Indonesia Based on K-Means Method"*
[75] *"Improving Foreign Language Proficiency in Society by Decision Tree Classification"*
[76] *"Implementation of the k-means clustering method for grouping poverty data: Case studies in Indonesia"*

A. RQ1 Distribusi Penggunaan Teknik *Data Mining* (2015-2024)

Dalam satu dekade terakhir, teknik *data mining* yang paling banyak diterapkan di Indonesia adalah *K-Means*, *Naive Bayes*, dan *Decision Tree*. Berdasarkan data yang terkumpul, *k-means* merupakan teknik yang paling umum digunakan, dengan 13 penerapan dalam berbagai publikasi.

Table 4 Teknik *Data Mining* di Indonesia

Technique	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	Total
K-Means			1	2	3	1	2		2	2	13
Naive Bayes		3			1	3	1	3	1		12
Decision Tree		2	3		1	2		1	2	1	12
K-Medoids			1			2	3	1			7
K-Nearest Neighbor					1		1	2		2	6
Support Vector Machine	1	1	1					2	1		6
Neural Network			1			1	1		1	1	5
Apriori			1			1	1				3
Deep Learning								1		1	2
Fuzzy C-Means	1		1								2
Relation of Rough Set				1							1
FP-Growth									1		1
Total	2	6	9	3	6	10	9	10	8	7	70

Teknik-teknik tersebut sering digunakan dalam tugas *clustering*, salah satu contohnya adalah studi oleh Prihandoko dalam jurnal *"Implementation of Data Mining for Determining Majors Using K-Means Algorithm in Students of SMA Negeri 1 Pangkalan Kerinci"* yang menunjukkan bagaimana *k-means* diterapkan dalam pemilihan jurusan siswa [24]. Selain itu, *Naive Bayes*, dan *Decision Tree* juga sangat populer, masing-masing dengan 12 penerapan. Teknik *Naive Bayes* banyak digunakan untuk klasifikasi, terutama di sektor kesehatan dan pendidikan. Misalnya, penelitian oleh Kurniawati dalam jurnal *"Comparative Study on Data Mining Classification Methods for Cervical Cancer Prediction Using Pap Smear Results"* menunjukkan aplikasi *naive bayes* dalam deteksi kanker serviks.

Tabel 5 Teknik dan Metode Data Mining yang Digunakan dalam Penelitian

PICO	Algorithm	Author
Association	Apriori	[18], [44], [48]
	Relation of Rough Set	[31]
	FP-Growth	[64]
Classification	Naive Bayes	[14], [16], [21], [36], [41], [45], [46], [55], [59], [62], [65], [73]
	Decision Tree	[15], [16], [22], [23], [25], [36], [41], [45], [62], [66], [68], [75]
	K-Nearest Neighbor	[35], [50], [58], [70], [72], [73]
Estimation & Forecasting	Support Vector Machine	[13], [20], [21], [63], [68], [73]
	Neural Network	[20], [41], [56], [65], [72]
	Deep Learning	[60], [71]
Clustering	K-Means	[24], [27], [28], [33], [36], [37], [42], [51], [57], [66], [67], [74], [76]
	K-Medoids	[24], [40], [47], [49], [50], [54], [61]
	Fuzzy C-Means	[17], [30]

Teknik *decision tree*, di sisi lain, diterapkan pada analisis data kesehatan, seperti yang dilakukan oleh Qudsi dalam jurnal *"Predictive Data Mining of Chronic Diseases Using Decision Tree: A Case Study of Health Insurance Company in Indonesia."* [21], [22].

B. RQ2 Analisis Tren Penggunaan di Berbagai Industri

Teknik data mining di Indonesia juga diaplikasikan di berbagai sektor industri, dimana tren teknik *data mining* di Indonesia memperlihatkan penggunaan yang signifikan dalam beberapa sektor utama, di antaranya adalah bisnis, kesehatan, dan pendidikan.

Tabel 6 *Data Mining* di Berbagai Sektor Industri

Teknik	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	Total
Business			2	1	2	2	2	1	3	2	15
Health		1	2		3	2	4	2	1		15
Education	1	3	1		1		2	1		2	11
Environment	1		2	1		3	1		1		9
Finance			1	1				1	1	1	5
Food						1		1	1	1	4
Governance				1			1	1			3
Technology		1						1			2
Total	2	5	8	4	6	8	10	8	7	6	64

Dalam sektor bisnis, data mining digunakan untuk memprediksi perilaku konsumen dan segmentasi pasar, sementara pada sektor kesehatan, fokusnya adalah pada prediksi dan deteksi penyakit serta pemetaan risiko kesehatan. Di bidang pendidikan, *data mining* diterapkan untuk menganalisis kinerja siswa dan mengidentifikasi faktor-faktor yang mempengaruhi hasil akademik.

Tabel 7 Sektor Industri Penggunaan *Data Mining*

Industry	Author
Business	[18], [19], [27], [32], [33], [39], [40], [48], [49], [62], [64], [65], [66], [70], [71]
Finance	[20], [28], [41], [58], [72]
Governance	[29], [57], [63]
Health	[21], [22], [30], [34], [35], [36], [42], [43], [50], [51], [52], [53], [59], [60], [67]
Environment	[13], [23], [24], [31], [44], [45], [46], [54], [68]
Food	[47], [69], [73], [74]
Education	[14], [15], [16], [17], [25], [37], [55], [56], [61], [75], [76]
Technology	[26], [62]

Data mining dalam bisnis banyak difokuskan pada pemahaman perilaku konsumen, prediksi churn, dan deteksi penipuan. Misalnya, algoritma seperti J48 dan Random Forest digunakan untuk memprediksi perilaku yang melibatkan penyalahgunaan promo dalam *e-commerce*. Teknik asosiasi seperti Apriori dan FP-Growth membantu mengidentifikasi pola pembelian, yang penting untuk segmentasi pasar dan pengelolaan inventaris. Misalnya, Supriyono dalam jurnal *"Implementation of Data Mining with Apriori Techniques to Determine the Pattern of Purchasing of Agricultural Equipment"* menerapkan teknik *association rules* untuk mengidentifikasi pola pembelian yang sering terjadi, yang bermanfaat bagi pengelolaan inventaris dan strategi pemasaran. Begitu juga Utami & Surjandari dalam jurnal *"Identifying Consumer Buying Behavior Differences Through Market Basket Analysis in Multiple Outlet Types"* menemukan bahwa *association rules* dapat meningkatkan prediksi kebutuhan pelanggan berdasarkan pola belanja [18]-[48].

Banyak penelitian dalam sektor kesehatan yang berfokus pada penggunaan *data mining* untuk mendeteksi dan memprediksi penyakit. Algoritma seperti Naive Bayes dan Decision Tree sering digunakan untuk memprediksi hasil kesehatan berdasarkan data pasien, seperti riwayat penyakit, gejala, dan hasil lab. Misalnya, prediksi penyakit diabetes atau deteksi dini kanker sering kali melibatkan algoritma klasifikasi untuk memisahkan data pasien ke dalam kategori risiko tertentu, sebagaimana ditunjukkan dalam studi oleh Fiarni dalam jurnal *"Analysis and Prediction of Diabetes Complication Disease using Data Mining Algorithm."* Teknik ini efektif dalam membantu dokter dan praktisi medis untuk melakukan deteksi dini dan intervensi yang lebih akurat [36].

Di sektor pendidikan, data mining berperan dalam mengevaluasi kinerja siswa, prediksi dropout, dan segmentasi kemampuan siswa. Teknik KNN, sebagai contoh, digunakan untuk memprediksi keberhasilan akademik berdasarkan faktor-faktor seperti prestasi sebelumnya dan tingkat motivasi siswa. Misalnya, Nasrullah dalam jurnal *"Predicting Student's Failure in Education Based on Dropout Status"* menggunakan teknik *classification* untuk memetakan faktor-faktor yang dapat mempengaruhi keberhasilan akademik, yang kemudian membantu institusi pendidikan dalam memberikan dukungan belajar yang lebih tepat sasaran [56].

5. SIMPULAN DAN SARAN

Kajian ini menyimpulkan bahwa *data mining* telah mengalami kemajuan yang signifikan di Indonesia dalam dekade terakhir, dengan penerapan yang signifikan di sektor bisnis, kesehatan, dan pendidikan. Hasil penelitian menunjukkan bahwa teknik seperti *K-Means*, *Naive Bayes*, dan *Decision Tree* paling sering digunakan. Di sektor bisnis, *clustering* melalui *K-Means* banyak diterapkan untuk segmentasi pasar dan analisis pola konsumen. Sektor kesehatan terutama memanfaatkan teknik *classification*, seperti *Naive Bayes* dan *Decision Tree*, untuk prediksi risiko penyakit dan diagnosis awal. Sementara itu, sektor pendidikan menggunakan *data mining* untuk menilai kinerja siswa dan memprediksi potensi *dropout*, membantu institusi dalam mengoptimalkan strategi pembelajaran.

Penelitian ini memiliki beberapa batasan yang perlu diperhatikan. Pertama, kajian literatur ini terbatas pada publikasi yang tersedia dalam database akademik terindeks SCOPUS, yang mungkin mengesampingkan beberapa studi lokal atau praktik yang tidak dipublikasikan secara luas. Kedua, penelitian ini berfokus pada tren penggunaan *data mining* di Indonesia tanpa mengeksplorasi secara mendalam faktor-faktor yang mempengaruhi adopsi teknik-teknik tertentu di tiap sektor. Dengan demikian, penelitian selanjutnya disarankan untuk mengeksplorasi tantangan dan faktor-faktor yang memengaruhi adopsi *data mining* di berbagai sektor, yang dapat memberikan wawasan lebih mendalam tentang kendala dan peluang dalam penerapan teknik ini di tingkat lokal. Juga, kajian lebih lanjut yang mencakup sumber data non-akademik atau laporan industri dapat membantu memperkaya pemahaman mengenai praktik-praktik *data mining* yang berkembang di Indonesia.

DAFTAR PUSTAKA

[1] R. Wijianto and Y. Nurdeviyanti, "Aplikasi Dewisita Pesona Wisata Di Jawa Tengah Berbasis Android," *J. Informatics Comput. Eng. J.*, vol. 3, no. 2, pp. 125–132, 2023.

[2] S. Sudaryanto, "Pengurangan Resiko Bencana Alam Melalui Pengembangan Kompetensi Aparatur Sipil Negara Badan Penangulangan Bencana Daerah di Jawa Tengah," *J. Public Policy Appl.*, vol. 3, no. 1, pp. 60–82, 2021.

[3] BNPB, "Kajian Risiko Bencana Jawa Tengah 2016 - 2020," *Badan Nas. Penanggulangan Bencana*, p. 68, 2020.

[4] R. Isnaini, "Analisis Bencana Tanah Longsor di Wilayah Provinsi Jawa Tengah," *Islam. Manag. Empower. J.*, vol. 1, no. 2, pp. 143–160, 2019, doi: 10.18326/imej.v1i2.143-160.

[5] D. Prabowo and R. Ayu Permatasari Rahman, "Identifikasi Potensi Banjir Di Kawasan Sub Urban Kota Semarang (Studi Kasus Kab Kendal)," *J. Planol.*, vol. 20, no. 1, pp. 1829–9172, 2023.

[6] D. A. Wicaksono and Y. A. Susetyo, "Clustering Zonasi Daerah Rawan Bencana Alam Di Provinsi Sumatera Barat Menggunakan Algoritma K-Means Dan Library Geopandas," *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 2, pp. 426–438, 2023, doi: 10.35870/jimik.v4i2.225.

[7] S. U. Darmatazia, D. P. Setianingrum, and M. P. Masyudi, "Analisis Penyebab Tanah Longsor di Gunung Manggah, Jalan Otto Iskandardinata, Kelurahan Sungai Dama, Kecamatan Samarinda Ilir, Kota Samarinda," *Dep. Ris. Dan Penelit.*, vol. 2, no. 1, pp. 1–9, 2022.

[8] F. A. Kurnianto, V. Elfiani, and A. F. Alfani, "Analisis Spasial Kerentanan Banjir dan Longsor di Kabupaten Banyuwangi," *JPIG (Jurnal Pendidik. dan Ilmu Geogr.*, vol. 6, no. 1, pp. 49–60, 2021, doi: 10.21067/jpig.v6i1.5323.

[9] B. N. P. Bencana, "Geoportal Data Bencana Indonesia," *Badan Nasional Penanggulangan Bencana*, 2023. https://gis.bnpb.go.id/ (accessed Nov. 26, 2023).

[10] O. sari Pinda rini and R. Khoirudin, "Analisis Sektor-Sektor Ekonomi Dengan Potensi Unggulan Di Kabupaten/Kota Se-Provinsi Jawa Tengah Tahun 2011-2016," *J. Ilm. Ekon. Dan Bisnis*, vol. 17, no. 1, pp. 15–23, 2020, doi: 10.31849/jieb.v17i1.2790.

[11] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 15, pp. 574–583, 2022, [Online]. Available: https://doi.org/10.5281/zenodo.7052276.

[12] N. L. Erawati, I. K. N. Adi Jaya, and I. K. A. Asmarajaya, "Visualisasi Data Pemetaan Nasabah Kredit Pada LPD Desa Adat Kesiman Menggunakan Python," *RESI J. Ris. Sist. Inf.*, vol. 2, no. 1, pp. 129–138, 2023, doi: 10.32795/resi.v2i1.4266.

[13] M. Murdiaty, A. Angela, and C. Sylvia, "Pengelompokkan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 744, 2020, doi: 10.30865/mib.v4i3.2213.

[14] F. Afrianto and A. D. Hariyanto, "Clustering Dan Visualisasi Bentuk Kota Di Jawa Timur," *J. Plano Buana*, vol. 2, no. 2, pp. 57–67, 2022, doi: 10.36456/jpb.v2i2.5303.

[15] E. Umargono, J. E. Suseno, and V. G. S. K., "K-Means Clustering Optimization using the Elbow Method and Early Centroid Determination Based-on Mean and Median," vol. 474, no. Isstec 2019, pp. 234–240, 2020, doi: 10.5220/0009908402340240.

[16] H. Y. Heraldi, N. C. Aprilia, and H. Pratiwi, "Analisis Cluster Intensitas Kebencanaan di Indonesia Menggunakan Metode K-Means," *Indones. J. Appl. Stat.*, vol. 2, no. 2, p. 137, 2019, doi: 10.13057/ijas.v2i2.34911.

[17] V. A. Ekasetya and A. Jananto, "Klusterisasi Optimal Dengan Elbow Method Untuk Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Semarang," *J. Din. Inform.*, vol. 12, no. 1, pp. 20–28, 2020, doi: 10.35315/informatika.v12i1.8159.

[18] H. Humaira and R. Rasyidah, "Determining The Appropiate Cluster Number Using Elbow Method for K-Means Algorithm," no. January 2020, 2020, doi: 10.4108/eai.24-1-2018.2292388.

[19] D. Cournapeau, "StandardScaler," *scikit-learn developers*, 2010. https://scikit-learn.org/dev/modules/generated/sklearn.preprocessing.StandardScaler.html (accessed Nov. 15, 2024).

[20] H. Hernandez, "Probability Distribution and Bias of the Sample Standard Deviation," *ForsChem Res. Reports*, vol. 8, pp. 1–26, 2023, doi: 10.13140/RG.2.2.22144.51205.

[21] E. Priambodo, T. Prahasto, and A. Widodo, "Deteksi Dini Kerusakan Bearing Menggunakan Machine Learning Pendekatan Support Vector Regression (Svr)," *J. Tek. Mesin S-1*, vol. 11, no. 3, pp. 101–112, 2023.

[22] E. Oktaviana, *Clustering bencana alam di Indonesia menggunakan algoritma K-Means*. Surabaya, 2022.

[23] B. Suharjo, "Application of K-Means Cluster and Spatial Statistics using Python to Analyze the Indicators of Indonesia Information Technology," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 12, no. 1, pp. 11–18, 2021, doi: 10.31849/digitalzone.v12i1.4310.

[24] M. R. N. Yusuf, Y. A. Pranoto, and F. X. Ariwibisono, "Rancang Bangun Sistem Informasi Geografis Pemetaan Upah Minimum Kota (Umk) Dan Biaya Kebutuhan Hidup Di Provinsi Jawa Timur Berbasis Web," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 4, no. 2, pp. 247–253, 2020, doi: 10.36040/jati.v4i2.2722.

[25] M. F. H. Adi and Y. A. Susetyo, "Pemetaan Potensi Bencana di Jawa Tengah Menggunakan Google Maps API dan KML dengan Metode," *Indones. J. Comput. Model.*, vol. 3, no. 1, pp. 43–51, 2018.