

Systematic Literature Review : Data Mining Algorithm for Decision Making

Subekti Prayogi^{1,*}, Hafizh Umar Syafiqh²

¹²Department of Computer Science
Esa Unggul University
Arjuna Utara No.9, Kebon Jeruk, Jakarta 11510

Received March 2025; accepted April 2025

ABSTRACT

Data mining is a technique used to explore and extract information hidden in big data, with the aim of discovering patterns or knowledge that can be used to support decision making. In recent years, data mining methods and algorithms have developed rapidly and applied in various fields, including education, business, technology, and e-commerce. This research aims to conduct a systematic review of methods and algorithms mining data used in decision making from 2016 to 2023. The method used in this research is Systematic Literature Review (SLR), which consists of three main stages: planning, conducting, and reporting. At the planning stage, the author determine the research theme and choose a Research Question (RQ) to limit the space scope and objectives of the study. Next, at the conducting stage, the author collected 25 journals relevant sources from various leading sources such as Elsevier and Springer. Results of this research is expected to provide a comprehensive overview of the development of methods and Algorithms used in decision making are based on data mining and trends the latest in the research. The research results will be presented systematically in the next chapter, with in-depth analysis regarding the application and effectiveness of the methods that have been used implemented.

Keywords: Systematic Literature Review; Data Mining; Decision Making

1. Introduction.

Data mining is a series of processes for exploring the added value of information from a database by extracting and recognizing what is important or interesting from the data [1-5]. Data Mining can be used to manage very large data with the aim of finding hidden information patterns [6]. Data mining is also very useful in scientific research using large databases. Data mining itself is very popular in various business sectors such as education, retail, technology, e-commerce, and others. This is because there is a need for an efficient data requirements analysis method to detect unknown information, such as technological development, predicting market interest and achieving goals for standardizing the vision and mission of a field. There are many different methods and algorithms in this research. This systematic review aims to identify and analyze data mining methods and algorithms used from 2016 to 2023. The method in this research will be explained in chapter 2, the results and answers to questions will be explained in chapter 3. Finally, the results of this research are summarized in the last chapter.

2. Methodology

The method used in this research uses a systematic literature review (SLR). The SLR method is used to analyze journal information that has been researched by previous researchers so that it can be reviewed systematically through several stages that must be prepared. The SLR

method uses 3 processes, namely Planning, Conducting and Reporting. Explanation of the 3 processes as follows:

1. Planning stage is the first step in determining the research theme. In this case the researcher chose the theme of data mining methods and algorithms for decision making. This theme is used by the author to determine objectives and select Research Questions (RQ). This Research Question (RQ) aims to limit the scope of research, guide research and evaluate research results. After the author determined the RQ, the author looked for journals related to the theme with the keywords Data mining methods and algorithms published from 2016 to 2023, as many as 25 journals from various sources such as Elsevier (sciencedirect.com), Springer (springerlink.com) or other journal publishing sources.
2. Conducting stage is the analysis of 25 journals, undertaken by the author, to identify the methods, algorithms, and sectors or fields in which data mining research was performed. The analysis is made in the form of an RQ table along with explanations as in table 1.

TABLE 1. Open Questionnaire

ID	Research Question	Main Motivation
RQ 1	What methods are most often used in the application of data mining?	Analyze what methods are most widely used in decision making using data mining.
RQ 2	What algorithms are most widely used in data mining?	Analyze what algorithms are most widely used in decision making using data mining.
RQ 3	What sectors or fields are often used in data mining research?	Analyze sectors or fields that are often used in research using data mining for decision making.

3. The reporting stage involves preparing a report that clearly and systematically explains the results of the review. In presenting the analysis results, the author explains descriptively the study findings and combines the data results using graphs with data percentages. The steps taken are to prepare the background and objectives of the research, carry out a journal data search process according to the chosen theme, create main results and carry out analysis and draw conclusions and recommendations from these results for the future.

3. Results and Discussion

3.1. What methods are most often used in the application of data mining?

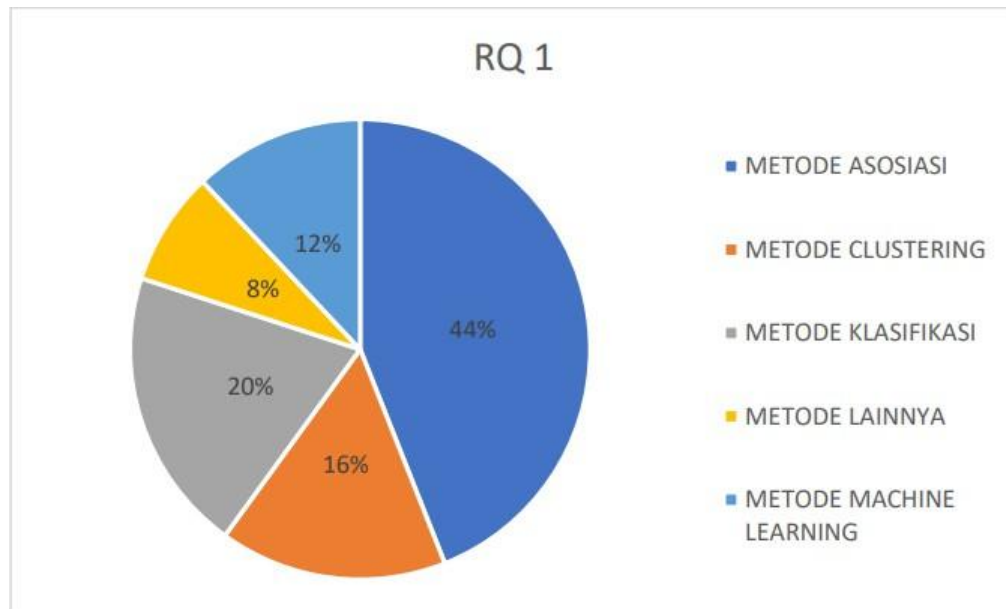


Figure 1. Data Mining Method

Based on Figure 3.1, the association method is the most frequently used with 11 times, indicating that analyzing the relationships between elements in a dataset (such as in purchasing or customer behavior analysis) is the main focus in this study.

Even though there are 2 times the use of other methods, the frequency is quite low compared to association, clustering, or classification methods. This may reflect the use of other, more case-specific techniques in the analysis.

The use of machine learning (3 times) shows that sophisticated methods that utilize automatic models for prediction or decision making are used in data analysis. While not as numerous as associations or classifications, it reflects the adoption of more modern techniques. Clustering (4 times) and classification (5 times) also received attention, reflecting the need to group data or predict categories based on data features.

3.2. What algorithms are most widely used in data mining?

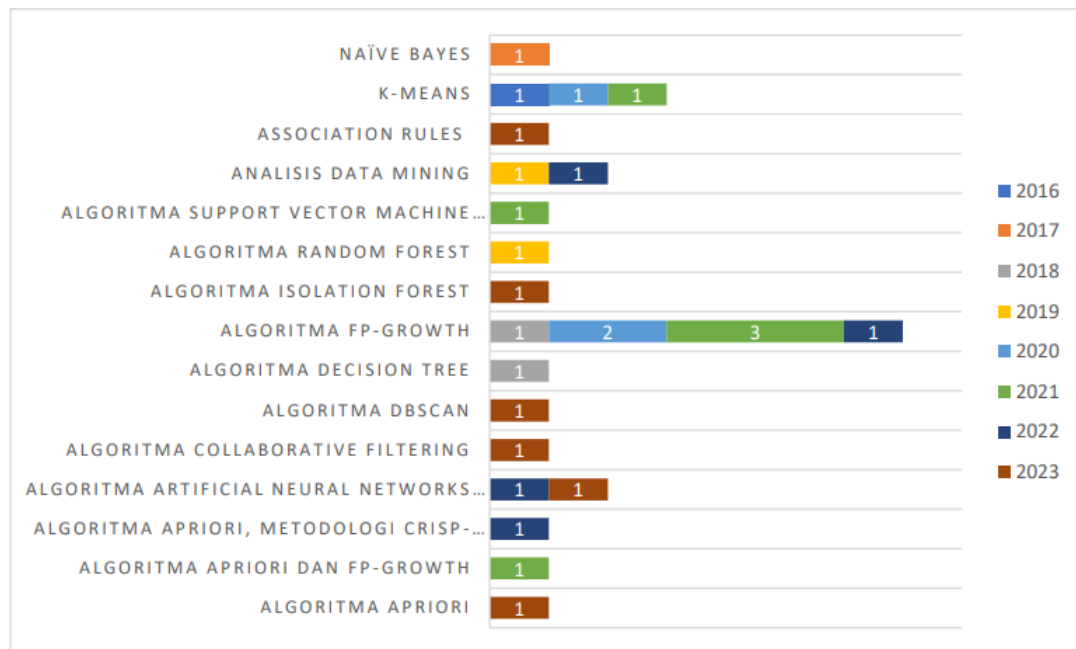


Figure 2. Number of Algorithm Uses by Year

Figure 3.2 explains that the FP-Growth algorithm has the highest number of uses (7 times) compared to other algorithms in this period. The use of this algorithm is distributed over several years, especially in 2018 (2 times), 2019 (3 times), and 2021 (1 time). This shows the popularity of the FP-Growth algorithm in associative pattern analysis and data mining.

A total of 15 different types of algorithms or methods were used during this period demonstrating the diversity of approaches to data analysis. Some algorithms that are only used once include Decision Tree, DBSCAN, Isolation Forest, and others.

Several algorithms have begun to be used in recent years, such as Artificial Neural Networks (ANN) and Isolation Forest in 2021 and 2023. This reflects the trend of adopting more complex and sophisticated algorithms in line with technological developments.

Algorithms such as K-Means (3 times) and Naïve Bayes (1 time) show a focus on clustering and classification analysis.

This indicates the importance of this method for certain applications in data mining. This data shows a trend of increasing use of diverse algorithms, with the FP-Growth algorithm being the most frequently used. The variety of algorithms reflects the increasingly extensive and complex needs of data mining analysis, with the introduction of advanced algorithms such as ANN in recent years.

3.3. What sectors or fields are often used in data mining research?

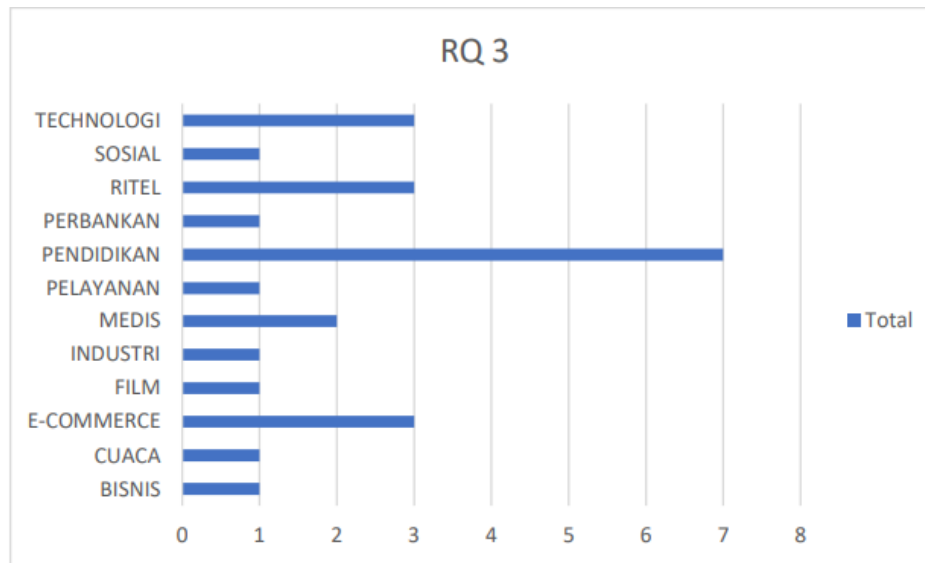


Figure 3. Field of Data Mining Research

The Education emerged as the category that used data mining the most with 7 times. This shows that the education sector relies heavily on data analysis techniques to improve learning experiences, assess student performance, and design better curricula [7-10].

E-commerce, Retail, and Technology each used data mining 3 times, indicating that these sectors also rely heavily on data to optimize operations, customer experience, and market trends [8-17].

Sectors such as weather, film, and social sectors have only used data mining once. This could indicate that although the application of data mining in these sectors may be limited, it remains relevant for some specific analyzes in these contexts [17, 19].

CNN Making an effective decision must be based on complex data analysis. Data mining is one of the main approaches that is often used in conducting research to obtain the latest information and is used in implementing decision making [20].

SLR identifies various data mining methods and algorithms that are used, namely Classification Algorithms, Clustering Algorithms, Association Rules mining, and Regression analysis [20-24].

Even though data mining has various advantages, there are several main challenges in its use, namely data quality, scalability and ethics and privacy.

4. Conclusions

Data mining is an important pillar in data-based decision making. Through this SLR, it is revealed that the success of implementing data mining algorithms is greatly influenced by choosing the right method, data quality, and understanding the domain. For the future, there needs to be tighter integration between data mining and advanced technologies such as IoT, cloud computing, and blockchain to increase the accuracy and speed of decision making.

REFERENCES

- [1] M. Abdulloh and A. Kusnawi, "Implementasi Data Mining Untuk Menemukan Pola Asosiatif Data Tracer Study," *Jurnal Teknologi dan Sistem Informasi*, vol. 10, no. 2, pp. 75- 85, 2018.
- [2] M. Efiyanna, S. Suryadi, and E. Prasetyo, "Tracer Study: The Alignment of Work Types With The Origin of Alumni Specialization of FKM UI," *Jurnal Pendidikan Kesehatan*, vol. 11, no. 3, pp. 102-112, 2019.
- [3] F. Wibowo and D. Jananto, "Implementasi Data Mining Metode Asosiasi Algoritma FPGrowth Pada Perusahaan Ritel," *Jurnal Sistem Informasi dan Teknologi*, vol. 14, no. 1, pp. 90- 100, 2020.
- [4] B. Budianto and S. Riana, "Penerapan Data Mining Menggunakan Algoritma FP-Growth Untuk Penentuan Strategi Promosi Fakultas Ilmu Komputer Universitas Kuningan," *Jurnal Teknik Informatika*, vol. 17, no. 2, pp. 55-65, 2020.
- [5] M. Syarif, H. Hamid, and A. Putra, "Algoritma FP-Growth Untuk Mengetahui Pola Perceraian Selama Masa Pandemi Covid-19 Di Kab. Majalengka," *Jurnal Statistik dan Komputasi*, vol. 18, no. 4, pp. 210-220, 2021.
- [6] D. Anggrawan, H. Siti, and F. Hidayati, "Menentukan Akurasi Tata Letak Barang dengan Menggunakan Algoritma Apriori dan Algoritma FP-Growth," *Jurnal Teknik Industri*, vol. 19, no. 3, pp. 72-82, 2021.
- [7] J. Patacsil and R. Acosta, "Analyzing The Relationship Between Information Technology Jobs Advertised On-line and Skills Requirements Using Association Rules," *International Journal of Data Science*, vol. 23, no. 2, pp. 45-55, 2021.
- [8] R. Tahir and S. Sitompul, "Penerapan Algoritma FP-Growth Dalam Menentukan Kecenderungan Mahasiswa Mengambil Mata Kuliah Pilihan," *Jurnal Pendidikan Teknologi Informasi*, vol. 12, no. 1, pp. 35-45, 2021.
- [9] D. Fitriani, T. Arini, and P. Setiawan, "Implementasi Association Rule Dengan Algoritma Apriori Pada Data Peminjaman Buku UPT Perpustakaan Universitas Lampung Menggunakan Metodologi CRISP-DM," *Jurnal Sistem Informasi Universitas Lampung*, vol. 20, no. 3, pp. 110-120, 2022.
- [10] M. Andika, E. Haryanto, and D. Prasetyo, "Penerapan Data Mining Pada Hasil Tracer Study Alumni Untuk Menemukan Pola Asosiasi Dengan Algoritma FP-Growth," *Jurnal Teknologi Pendidikan*, vol. 21, no. 2, pp. 55-65, 2022.
- [11] D. Subekti, I. Wibowo, and A. Yulianto, "Aplikasi Data Mining Untuk Menganalisis Aturan Asosiasi Tracer Study Di Ibi-Kosgoro1957," *Jurnal Teknik Komputer*, vol. 18, no. 1, pp. 98-108, 2023.
- [12] D. Purnomo and I. Maulana, "Penerapan Algoritma K-Means untuk Clustering Data Transaksi E-Commerce di Indonesia," *Jurnal Data Mining dan Analisis*, vol. 9, no. 1, pp. 45- 58, 2016.
- [13] A. Setiawan and S. Aisyah, "Analisis Sentimen Ulasan Produk Menggunakan Algoritma Naïve Bayes," *Jurnal Teknologi Informasi*, vol. 10, no. 2, pp. 23-30, 2017.
- [14] D. Pratiwi and B. Santoso, "Penerapan Data Mining untuk Prediksi Penyakit Jantung di Indonesia," *Jurnal Kesehatan Masyarakat*, vol. 16, no. 1, pp. 50-59, 2018.
- [15] F. Ibrahim and L. Handayani, "Penerapan K-Nearest Neighbors (KNN) untuk Prediksi Kegagalan Mesin Industri," *Jurnal Teknik Industri*, vol. 19, no. 2, pp. 115-126, 2020.
- [16] R. Suryanto and E. Wulandari, "Prediksi Harga Saham Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Ekonomi dan Bisnis*, vol. 22, no. 3, pp. 97-106, 2021.
- [17] D. Setiawan and L. Salim, "Penerapan Clustering pada Data Pengunjung Situs Web untuk Rekomendasi Konten," *Jurnal Sistem Informasi dan Teknologi*, vol. 14, no. 4, pp. 135-145, 2021.
- [18] Y. Anggraini and N. Riza, "Pemanfaatan Data Mining untuk Analisis Keluhan Pelanggan di Layanan Pelanggan," *Jurnal Manajemen dan Pemasaran*, vol. 19, no. 2, pp. 75-84, 2022.
- [19] S. Prasetyo and A. Amiruddin, "Penggunaan Algoritma Artificial Neural Networks (ANN) untuk Prediksi Penjualan Produk Retail," *Jurnal Teknologi dan Sistem Informasi Bisnis*, vol. 16, no. 1, pp. 50-60, 2022.
- [20] A. Kurniawan and D. Wulandari, "Optimalisasi Rekomendasi Film Menggunakan Collaborative Filtering Berbasis Data Mining," *Jurnal Komputasi dan Informatika*, vol. 24, no. 3, pp. 123-134, 2023.

- [21] R. Setyawati and F. Mahendra, "Deteksi Anomali dalam Lalu Lintas Jaringan Menggunakan Algoritma Isolation Forest," *Jurnal Teknologi Keamanan Jaringan*, vol. 25, no. 2, pp. 110-120, 2023.
- [22] A. Nugroho and D. Yuliana, "Clustering Data Penyakit Menggunakan Algoritma DBSCAN untuk Identifikasi Jenis Penyakit," *Jurnal Data dan Analisis Kesehatan*, vol. 19, no. 1, pp. 65-75, 2023.
- [23] M. D. Santoso and A. Kurniawan, "Implementasi Data Mining dalam Analisis Perilaku Konsumen di Toko Online," *Jurnal Teknologi E-Commerce*, vol. 20, no. 3, pp. 150-160, 2023.
- [24] M. D. Santoso and A. Kurniawan, "Implementasi Data Mining dalam Analisis Perilaku Konsumen di Toko Online," *Jurnal Teknologi E-Commerce*, vol. 20, no. 3, pp. 150-160, 2023.