



A Review Of Data Mining Techniques And Their Role In Strategic

¹ Dr. Sushilkumar.R. Kalmegh ,² Anup S. Bele,
¹ Professor, ² Research Scholar,
¹ Computer Science Department

¹ Sant Gadge Baba Amravati University, Amravati, India,² Sant Gadge Baba Amravati University, Amravati, India

Abstract: Now in real life some of interesting role of human life are data collection, information, knowledge. Data mining is process to discover hidden knowledge from huge amount of dataset. In the last few years most of the techniques and application survey through the research. It used in multiple areas including education, finance, healthcare, industry, agricultural and cyber security. Data mining uses statistics, machine learning, artificial intelligence and visualization techniques to decision making. This review explorer the requirement and challenges about the data mining techniques and tools. Data mining techniques have become essential tools for research pattern recognition, trends, forecasting from vast dataset across various domains. various types of data mining tools are available like MINERWEKA,SAS,RAPID,KNIME,ORANGE,RLANG, MATLAB, PYTHON. This study collects summary of information about the basic concept of data mining and techniques used for decision making in distinct domain.

Index Terms - Data Mining, Data Mining tools, Data Mining Process, Data Mining Application.

I. INTRODUCTION

Data Mining is the process to manipulate large amount of dataset for the solving distinct problem in different field. Techniques are the particular way of doing activities. A key data mining techniques is available to take valuable decision. classification, association, clustering, decision trees. It used, prediction and regression. A key method for locating hidden patterns, correlations, and trends in big datasets is data mining [1].Improvement of data mining techniques by using the Machine Learning, Artificial Intelligences and execute modern techniques of previous decade. It offers a thorough approach to information discovery by combining techniques from database systems, statistics, and machine learning[2].Tools are computer programs to assist developer. Interactive data mining tools are available for data visualization and data analysis, which are design for improvement of processing. Explain mining process and tools Domain-driven data mining, which emphasizes the value of incorporating domain expertise into the mining process, has become a viable tool as organizations deal with more complex data-driven difficulties[3]. Elaborate different tool the effectiveness of decision-making is further increased by interactive data visualization and exploratory analysis tools, which are made possible by improvements in processing and storage capacities[4].This paper highlights techniques and tools used in various fields, offering insights into how domain-specific challenges are addressed.

II ,Data Mining Techniques

When we refer to data mining, we mean the process of finding potentially useful patterns by using big data. During this process, machine learning, statistics, and AI is used to extract information in a distinct domain. The data mining comprises data classification, data transformation, data integration, data discretization and pattern evaluation and more. Data mining techniques are used by the organizations to solve different business problems.

3.1 Classification

Classification is a process that assigns data to predefined classes. It is widely used across various fields and relies on several common algorithms, including Decision Tree (C4.5), Naive Bayes, and Support Vector Machine (SVM). In education, classification techniques are applied to predict student performance [11]. In the finance sector, they help in loan default prediction [12]. Healthcare benefits from classification by enabling accurate disease diagnosis, such as in diabetes classification [11]. In industry, it assists in quality inspection tasks, distinguishing between defect and non-defect products [12]. In agriculture, classification is used for crop disease identification [12], while in cyber security, it plays a critical role in detecting intrusions and classifying malware [12].

3.2 Clustering

Clustering is a technique that groups similar data without the use of predefined labels. It employs common algorithms such as K-Means, DBSCAN, and Hierarchical Clustering. In education, clustering is used to group students based on their learning patterns [11]. In finance, it aids in customer segmentation for targeted marketing strategies [12]. In healthcare, clustering helps group patients for risk stratification and better management of care [11]. Within the industrial sector, it is applied to group machinery behavior for maintenance and optimization purposes [12]. In agriculture, clustering assists in soil or region segmentation to enhance farming practices [12]. In cyber security, it plays an important role by grouping similar attack behaviors to detect and respond to threats effectively [12].

3.3 Association Rule Mining

Association is a technique used to find relationships among variables. Common algorithms for association include Apriori and FP-Growth. In education, association techniques help in discovering patterns in learning behaviors [11]. In finance, they are widely used for market basket analysis to identify products that are frequently bought together [12]. In healthcare, association methods assist in uncovering drug interactions that could impact patient care [11]. In the industrial sector, they reveal correlations between different types of machinery failures [12]. In agriculture, association techniques link environmental conditions to crop success, helping optimize farming practices [12]. In cyber security, they are crucial for detecting frequent sequences of threats and attacks [12].

3.4 Regression

Regression analysis is a technique used to predict numerical outcomes. Common algorithms in this category include Linear Regression and Ridge/Lasso Regression. Regression methods are widely applied across different sectors. In education, they are used to forecast GPA or academic scores [11]. In finance, they assist in stock market prediction [12]. In healthcare, regression models help predict patient stay durations or recovery times [11]. In the industrial sector, they are used for predictive maintenance modeling [12]. In agriculture, regression techniques support yield prediction based on weather and soil data [12]. In cybersecurity, they help in predicting the probability of breaches [12].

3.5 Anomaly Detection

Anomaly detection focuses on identifying unusual or suspicious data points. Common algorithms used for anomaly detection include Isolation Forest, Local Outlier Factor (LOF), and Auto encoders. This technique is applied across a variety of fields. In education, it is used to detect unusual academic activities [11]. In finance, anomaly detection helps in identifying fraudulent transactions [12]. In healthcare, it aids in spotting abnormal lab results [11]. In industry, it plays a crucial role in early failure detection in machines [12]. In agriculture, it assists in detecting unusual pest or disease outbreaks [12]. In cyber security, anomaly detection is essential for building intrusion detection systems (IDS) [12].

3.6 Sequential Pattern Mining

Sequential pattern mining focuses on extracting frequently occurring sequences of events or behaviors. Common algorithms for this technique include GSP (Generalized Sequential Pattern) and Prefix Span. Sequential pattern mining is widely applied across different sectors. In education, it is used to identify learning behavior trends over time [11]. In finance, it helps analyze purchasing or investment sequences [12]. In healthcare, it assists in monitoring patient disease progression [11]. In industry, it supports the discovery of operational event patterns [12]. In agriculture, it reveals seasonal crop activity patterns [12]. In cyber security, it is utilized to analyze threat evolution and propagation [12].

4.0 Data Mining Tools

4.1 WEKA

WEKA (Waikato Environment for Knowledge Analysis) is an open-source suite supporting various machine learning tasks, including data preprocessing, classification, regression, clustering, and visualization. It is known for its simplicity and effectiveness in academic and practical environments [2].

4.2 Rapid Miner

Rapid Miner provides a unified platform for data preparation, machine learning, and deep analytics. It supports a drag-and-drop interface, allowing users to design workflows easily, and is widely used in commercial and research applications [5].

4.3 KNIME

KNIME offers a modular data pipelining environment that integrates various components for machine learning and data analytics. It enables non-programmers to perform complex tasks through a graphical interface and has been widely used in life sciences and financial analytics [6].

4.4 Orange

Orange is a data mining tool built in Python with a visual programming interface. It supports interactive data visualization and modeling and is especially useful in education and research [7].

4.5 SAS Enterprise Miner

SAS is a commercial analytics suite offering extensive data mining capabilities, including decision trees, regression, and neural networks. It is particularly prevalent in enterprise applications for fraud detection, forecasting, and customer analytics [8].

4.6 R Language

R is a statistical programming language with a rich ecosystem for data analysis and mining. With thousands of user-contributed packages, it supports advanced statistical modeling and visualization [9].

4.7 Python

Python, with libraries like scikit-learn, pandas, and TensorFlow, is one of the most versatile languages for data science. It enables scalable and efficient implementation of data mining tasks in both research and production environments [10].

5.0 Data Mining Process

In the data mining process, the first step is data collection. Different sources such as databases, the web, IoT devices, sensors, data warehouses, and other data repositories contain a huge amount of data. The second step is data preprocessing, which uses techniques such as data cleaning, data integration, and data transformation, often with the help of various data mining tools. The third step is data selection, where relevant data attributes are chosen for further analysis. The fourth step is pattern evaluation, which involves evaluating interesting patterns and removing redundant information. The fifth step is knowledge representation, where the insights are displayed, often using dashboards, to support decision-making.

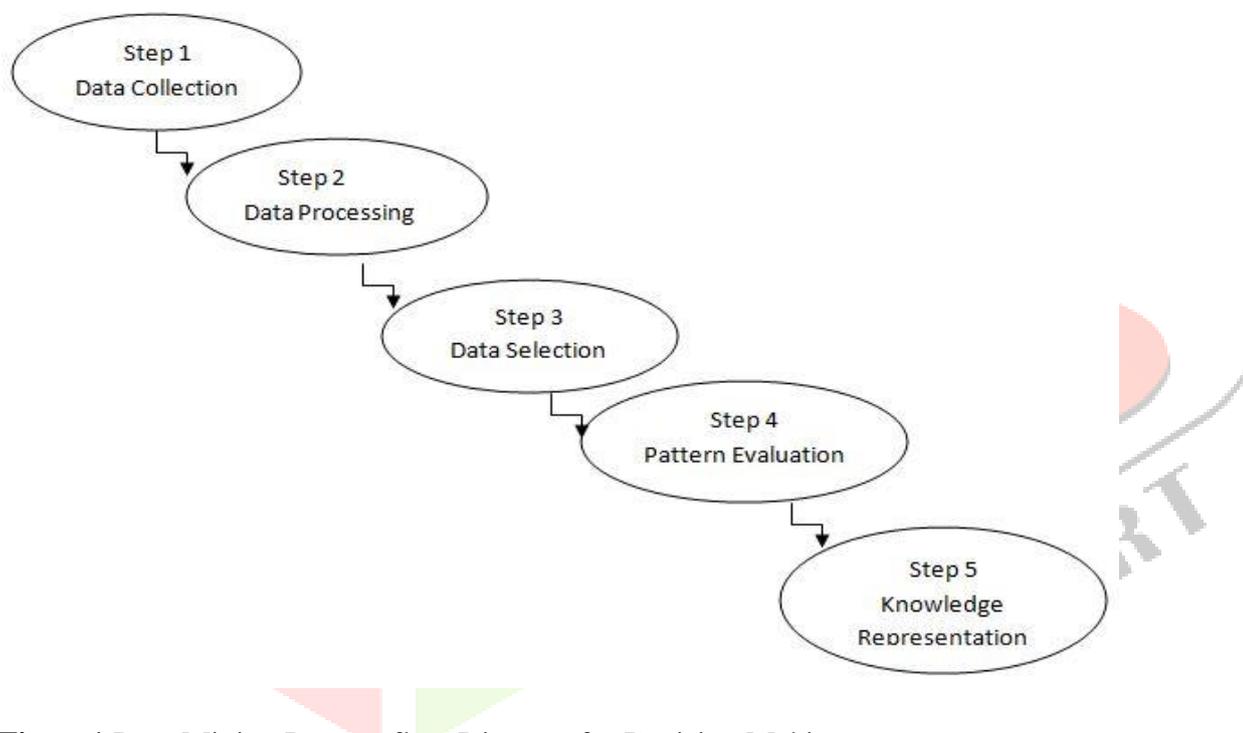


Figure1.Data Mining Process flow Diagram for Decision Making.

Table 6: classification for Domain, Data Mining Techniques and Decision-Making application [11],[12].

Domain	Data Mining Technique	Decision-Making Application
Healthcare	Classification, Clustering	Diagnosing diseases, predicting patient outcomes
Finance	Association Rule Mining, Prediction	Fraud detection, credit scoring, risk assessment
Marketing	Clustering, Association Rules	Customer segmentation, targeted advertising
Manufacturing	Anomaly Detection, Classification	Predictive maintenance, quality control
Retail	Association Rule Mining, Clustering	Market basket analysis, inventory management
Agriculture	Classification, Regression	Crop yield prediction, disease detection
E-commerce	Recommendation Systems, Clustering	Product recommendations, customer behavior analysis
Energy Sector	Time Series Analysis, Forecasting	Energy consumption prediction, grid optimization
Telecommunications	Clustering, Outlier Detection	Customer churn analysis, network fault detection
Education	Classification, Clustering	Student performance prediction, dropout prevention

7.0 Conclusion

This paper covers the Data Mining Techniques and its tools for decision-making in distinct domain. In today's world each and every person is using data mining knowingly and unknowingly. Also this paper focus on tools of data mining WEKA, SAS, RAPID MINER, KNIME, ORANGE, RLANG, MATLAB, PYTHON used for most of the Data Mining tasks. As a future prospective work, we are going to study various data mining Techniques and its tools for decision-making in distinct domain.

References

- [1] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, no. 3, pp. 37–54, 1996.
- [2] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed., Morgan Kaufmann, 2016.
- [3] C. Zhang, P. S. Yu, and D. Bell, "Domain-driven data mining: A practical methodology," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 9, pp. 1299–1312, Sept. 2010.
- [4] J. Gryz, "Interactive visualization of large data sets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 8, pp. 2142–2157, Aug. 2016.
- [5] M. Hofmann and R. Klinkenberg, "RapidMiner: Data Mining Use Cases and Business Analytics Applications," *CRC Press*, 2013.
- [6] M. R. Berthold et al., "KNIME: The Konstanz Information Miner," in *Proc. 4th Int. Industrial Conf. Data Mining*, Leipzig, Germany, 2004, pp. 319–326.
- [7] J. Demšar, B. Zupan, and G. Leban, "Orange: From Experimental Machine Learning to Interactive Data Mining," *White Paper*, Faculty of Computer and Information Science, University of Ljubljana, 2004.
- [8] SAS Institute Inc., *SAS Enterprise Miner 14.1: User's Guide*, Cary, NC, USA: SAS Institute, 2015.
- [9] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020.
- [10] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
- [11] S. Kaur and S. Gill, "A Survey on Data Mining Techniques in Healthcare:"

Opportunities and Challenges," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 874-8
DOI: 10.1109/Confluence51648.2021.9377148

[12] A. A. Laghari, K. Ahmed, M. S. Arshad, R. Shaikh, and M. A. Laghari, "Data Mining Techniques: A Survey Paper," 2017 9th IEEE-GCC Conference and Exhibition (GCCCE), Manama, Bahrain, 2017, pp. 1-6. DOI: 10.1109/IEEEGCC.2017.8448140

